Combining Low-dimensional Models with High-fidelity Data: A Multi-fidelity Approach to Transient Heat Transfer Problems

by

Pengdi Zhang

Bachelor's degree, University of Shanghai for Science and Technology, 2018

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Pengdi Zhang

It was defended on

April 5, 2021

and approved by

Peyman Givi, PhD, Professor, Department of Mechanical Engineering and Materials Science

Shervin Sammak, PhD, Assistant Professor, Department of Mechanical Engineering and Materials Science

Thesis Advisor: Hessam Babaee, PhD, Assistant Professor, Department of Mechanical Engineering and Materials Science

Combining Low-dimensional Models with High-fidelity Data: A Multi-fidelity Approach to Transient Heat Transfer Problems

Pengdi Zhang, M.S.

University of Pittsburgh, 2021

We present a generic multi-fidelity approach for combining constructing multi-fidelity surrogate models for transient heat transfer applications. We apply our developed methodology to build various surrogate models for mixing temperature of two different fluids. This is a classical heat transfer problem with numerous applications in diverse industries and it is considered in this work as a template problem for our methodology. In the presented framework, data from various levels of fidelity can be combined in a principled manner. More broadly, our target applications are problems where relying on high-fidelity data alone is not sufficient to build satisfactory surrogate models due to the high expense often associated with high-fidelity data acquisition. On the other hand, it may be possible to build reduced order models that can be sampled at a high rate with insignificant computational cost. However, the ROM may be inaccurate due to physics deficiency of the full-dimensional model as well as the reduction errors. To this end, we utilize reduced-order models of heat transfer mixing, i.e. low fidelity model, and high-fidelity measurements, which are obtained by performing direct numerical simulations. We will combine these two data sources using Gaussian process regression (GPR) and auto-regressive stochastic strategy. GPR is a non-parametric Bayesian regression technique that has a fully probabilistic workflow, in which the prediction uncertainties can be quantified in a principled manner. In the following research, we will successively verify the accuracy and computational effectiveness of multi-fidelity results for different quantity of parameters and different size of training data.

The novelty of the presented methodology is that: (i) Multi-fidelity Gaussian process regression; (ii) Spectral/hp element method; (iii) Proper orthogonal decomposition.

Table of Contents

Preface . . . xi

1.0 Introduction . . . 1

1.1 Numerical simulation for physical phenomenon . . . 1

1.2 Multi-fidelity modeling . . . 2

1.3 Dimension reduction . . . 4

2.0 Problem description and setup . . . 6

3.0 Methodology . . . 11

3.1 High-fidelity Model . . . 11

3.2 Low-fidelity Model . . . 13

3.2.1 Singular Value Decomposition . . . 14

3.2.2 The Least Square Regression . . . 16

3.2.3 Cross-Validation Test . . . 18

3.2.4 Proposed Approach . . . 19

3.3 Multi-Fidelity Modeling . . . 21

3.3.1 Multi-Fidelity Scheme . . . 23

3.3.1.1 Training . . . 24

3.3.1.2 Prediction . . . 26

4.0 Results . . . 27

4.1 Prediction Result With 1-D Multi-fidelity Framework . . . 27

4.1.1 The Cross-validation For Least Square Regression . . . 28

4.1.2 Temporal Multi-fidelity Modeling . . . 32

4.2 Prediction Result With 2-D Multi-fidelity Modeling . . . 36

4.2.1 The Cross-validation For Least Square Regression . . . 36

4.2.2 Space-time Temperature Multi-fidelity Modeling . . . 39

4.2.3 The Comparison With Single-fidelity Modeling . . . 43

5.0 Conclusions . . . 45

Bibliography 46

List of Figures

1 Upper left panel: the snapshot of high-fidelity numerical simulation; Lower left panel: the low-fidelity level data based on dimension reduction; Right panel: the prediction of quantity of interest obtained from numerical simulation results and low-dimensional data through Gaussian process regression. This study obtained a high-fidelity approximation from numerical simulation, and low-dimensional data is viewed as a low-fidelity approximation. 3

2 (i) Snapshots of temperature contours inside the elbow (time unit = 70). (ii) Snapshots of temperature contours insider the elbow (time unit = 72). (iii) Schematic of mixing elbow. From Fig (i) and (ii), the high-temperature fluid from the small inlet A is injected into the low-temperature fluid in the main pipe and mixed in the pipe. Heat transfer occurs at the elbow and the second half of the pipe. 7

3 Unstructured quadrilateral grid; (i) two-dimensional view of elbow mixture. (ii) x-y view of the grid in the nearness of pipeline elbow grid. The grid become denser at the junction of the vertical pipeline and the main pipeline. 8

4 Based on the existing numerical simulation data, our approach is split into three steps: 1)Upper left panel: Decompose and reduce the dimension of the existing numerical simulation data set A. 2)Upper right panel: Establish a low-fidelity model for a particular ξ^* by predicting the right-singular vector of the corresponding data $f(x, \xi^*)$. 3) Lower panel: For a given variable xi^*, combine its corresponding numerical simulation data f_{HF} and low-precision model f_{LF} to generate a multi-precision framework for flow temperature. 12

5 In present study, the linear regression model based on dimensional reduction is regarded as the heat transfer problem's low-accuracy estimation. At the same time, the high-fidelity observation data is driven by numerical simulation tool. The uncertainty that is specified by blue area shows two times standard deviation ($\pm 2\sigma[T]$) around the estimation work for the multi-fidelity models. 22

6 The study domain of multi-fidelity modeling. In $1-D$ multi-fidelity framework, the prediction target on the average temperature of outlet surface. 27

7 The L_2 mean error (i) and standard deviation (ii) of least square regression with the change of quantity of training points. We construct one hundred linear regression models by randomly sampling the original data set and get the error of L_2 by averaging all the models (Here, the rank of r equal to 3 and T_a related variables ξ and η are two vectors with the size of 5×1.). The low standard deviation indicates the reliability of the model for unobserved cases and the low training points quantity means low computational cost. 29

8 Temporal low-fidelity model: Correlation of average temperature obtained from low-fidelity (f_{LF}) with different quantity of training sets (i)$S_t = 8$, (ii)$S_t = 12$, (iii)$S_t = 16$, (iv)$S_t = 20$. The red line shows a perfect correlation between low–fidelity data and high-fidelity data. These figures show that as the quantity of training sets increases, the low-fidelity data points get closer to the straight line $x = y$. 31

9 Temporal multi-fidelity model: Correlation of average temperature obtained from (i) low-fidelity and (ii),(iii),(iv) multi-fidelity models with different quantity of high-fidelity data points (ii) $N_{HF} = 5$, (iii) $N_{HF} = 9$, (iv) $N_{HF} = 13$, where 201 low fidelity points are used. These plots show that as the quantity of high-fidelity observations increases, the multi-fidelity model becomes more accurate. 33

10 Temporal multi-fidelity model: Average temperature comparisons between multi-fidelity, high-/low-fidelity versus time t. In Fig.10, $N_{LF} = 201$ points are used and (ii), (iii) and (iv) show the effect of improving the quantity of high-fidelity data points, (ii) $N_{HF} = 5$, (iii) $N_{HF} = 9$, (iv) $N_{HF} = 13$. These plots show that as the quantity of high-fidelity observations increases, the multi-fidelity model turns to more accurate. 34

11 The L_2 mean error (i) and standard deviation (ii) of least square regression with the change of quantity of training points. We construct one hundred linear regression models by randomly sampling the original data set and get the error of L_2 by averaging all the models (Here, the rank of r equal to 3 and T_a related variables ξ and η are two vectors with the size of 5×1.). Compared with the result in previous section, the trends of mean error and standard deviation are similar. 36

12 Space-time low-fidelity model: Correlation of temperature obtained from (i),(ii),(iii),(iv) low-fidelity(f_{LF}) with different quantity of training points (i) $N_t = 10$, (ii) $N_t = 12$, (iii)$N_t = 14$, (iv) $N_t = 19$. These plots show that as the quantity of high-fidelity observations increases, the low-fidelity model becomes more accurate. 38

13 Space-time multi-fidelity model: Correlation of temperature obtained from (i) low-fidelity and (ii),(iii),(iv) multi-fidelity models with different quantity of training points (ii)$N_t = 25$, (iii)$N_t = 35$, (iv)$N_t = 45$. In the multi-fidelity modeling, $N_{LF} = 2121$ points are used. These plots show that as the quantity of high-fidelity data sets grows, the multi-fidelity model becomes more precise. 40

14 Space-time high-fidelity model and low-fidelity model: the contour for outflow surface temperature over time obtained from true observation and low-fidelity points. Since the temperature change at the right end of outlet C is not obvious, the low-fidelity model is slightly different from it. The high-fidelity data shows that the temperature change period and phase at the left end of outlet C is obviously different from the low-fidelity data. 41

15 Space-time multi-fidelity model: The contour for outflow surface temperature over time obtained from multi-fidelity models with different quantity of training points, (i)(ii)$N_t = 25$, (iii)(iv)$N_t = 35$, (v)(vi)$N_t = 45$. The Figure of uncertainty shows the $\pm 2\sigma[T]$(standard deviation) around the mean for the multi-fidelity models. 42

16 Space-time multi-fidelity model and single-fidelity model: The contour for outflow surface temperature over time obtained from singular-fidelity models and multi-fidelity models; the quantity of high-fidelity data points are both 45. Obviously, a large amount of low-precision data points has greatly improved the accuracy of the multi-fidelity model. 43

Preface

First of all, I want to thank my advisor Dr. Hessam Babaee. He is a very patient and kind person. During the remote conversation during the epidemic, his guidance gave me a deeper understanding of multi-precision modeling. His empathy and motivation supported me to complete my master's thesis remotely during the epidemic. At the same time, I would like to thank postdoc Donya and my friend Alireza Amiri Margavi for their valuable comments and encouragement on my thesis. In the simulation part, Phd Prerna Patil provided a lot of guidance so that I can carry out the direct numerical method by Nektar smoothly.

Thanks to my parents for their unconditional love and support for me. Without your care and protection, I would not be where I am today.

1.0 Introduction

In most industrial processes, people use pipeline mixing techniques to promote chemical reactions, heat transfer, mixing, and combustion processes. Here, complete knowledge of the principles that rule the phenomena involving fluids transportation leads to more efficient and secure systems. As discussed by Linda M. Sroka and Larry J. Forney(1989) [19], the conventional configuration of pipeline mixing is a side tee used to connect multiple pipelines followed by a straight pipe length. However, in many industrial sites, such as petroleum, chemical, oil, and gas industries, the pipeline's side tee flow is complicated. Meanwhile, after the mixing tee, the mixing process does not have enough distance to achieve the desired mixing quality. Here, the elbow in the pipeline near the injection point may generate sufficient secondary flow and turbulence intensity, thereby significantly shortening the pipeline length to achieve the desired degree of mixing.

Here there has been plenty of research on elbows ([1],[8],[5]). However, we still want to show the heat transfer process of fluid in elbows by analyzing the mixed flow of elbows to illustrate how the new analysis method, multi-fidelity modeling, solve traditional industrial problems with more obvious advantages. For this study, the cold flow of the main pipeline and the hot flow injected at the elbow cause the two fluids to mix the next pipe and generate heat transfer rapidly.

1.1 Numerical simulation for physical phenomenon

Since the Middle Ages, various physical phenomena have been essential and active research topics because they affect multiple applications such as aviation, transportation, energy generation systems, and weather forecasting. In particular, turbulence is one the most challenging problems in fluid dynamics. Commonly experimental or computational studies are conducted for a variety of predictive and diagnostic purposes. Although experiments can yield valuable and accurate data, experiments are often expensive and require many

human resources and time. Here, CFDs can be used as extra work to help people understand turbulence in more detail, especially in some cases where the experiment is too expensive or impossible to achieve. Although with the advancement of science and technology, high-fidelity CFD technology represented by numerical simulation can improve people's understanding of fluid phenomena. Another problem arises, numerical simulation calculation is very time-consuming and has high requirements for the computer's computing power. We call it computationally expensive. At present, how to reduce the demand for computing power while modeling the dynamics of fluid phenomena has become an active research topic. In this study, we take the elbow as an example to show how to model high-fidelity dynamics of elbow mixture while significantly minimizing the steep calculation and data storage costs associated with numerical simulation.

1.2 Multi-fidelity modeling

Surrogate models are widely used in many complex engineering systems for industrial scenarios. The surrogate models can be used for a variety of purposes, e.g., optimization, control, uncertainty quantification, system-level integration of components and many other applications that require fast-to-evaluate models. In this study, we introduce a method called multi-fidelity modeling. It uses the latest research in machine learning to complete the trade-off between computational efficiency and prediction accuracy. As shown in the Fig 4, for producing prediction of a typical quantity of interest (QoI), we will establish a high-fidelity model (numerical simulation) and a low-fidelity model (dimension reduction) at the same time and combine the predictions of these two models. These various sources of data have their advantages and disadvantages. For example, numerical simulation response or empirical correlations contain high-fidelity information, but they are very costly to obtain and therefore these high-fidelity predictions are often scarce. On the other hand, the predictions based on dimension reduction alone can lack sufficient accuracy, but once the reduced system is constructed, these predictions can be made with little cost. The overall objective of this thesis is to develop a multi-fidelity framework that can utilize the scarce but accurate high-

fidelity measurements with ROM predictions that may have inaccuracies, but they are cheap to evaluate and therefore we have plenty of these measurements. Therefore, we hope that most of the data in our multi-fidelity system consist of low-precision data. These low-precision data may not be accurate enough, but they can be collected quickly to improve the entire system's computational efficiency. Therefore, the current research goal is to merge the low-precision information in the low-latitude model with a small amount of high-fidelity data from the numerical simulation to solve the enormous challenge of weighing the trade-off between computing efficiency and prediction accuracy.

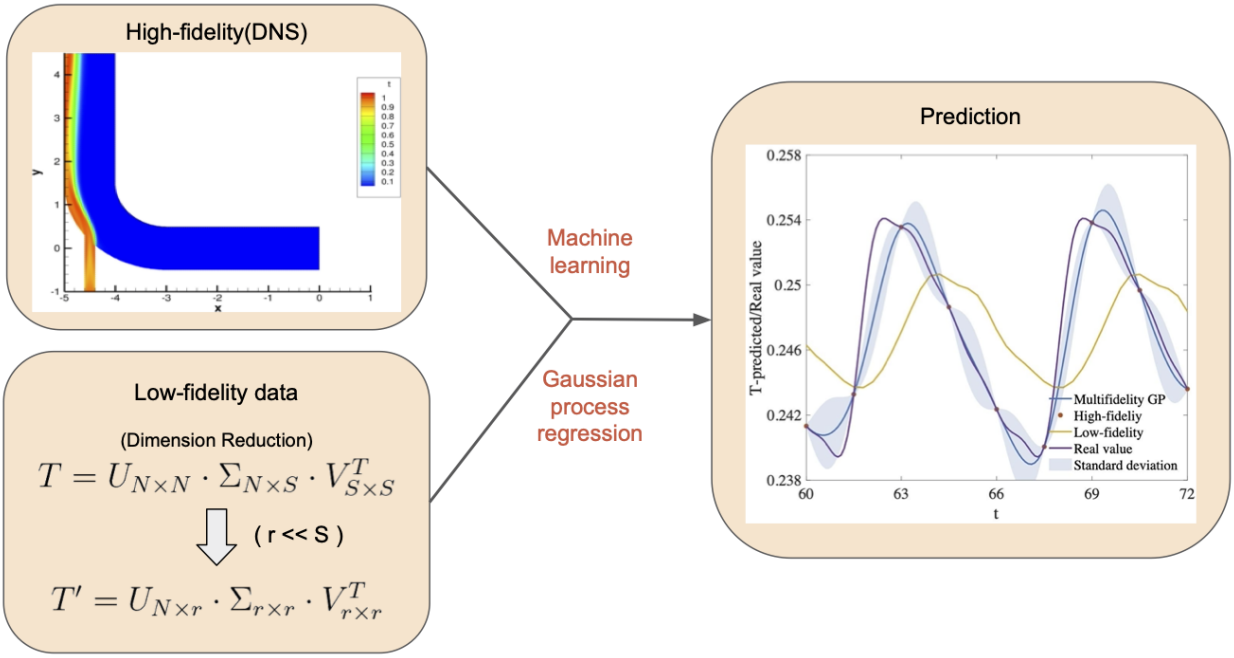


Figure 1. Upper left panel: the snapshot of high-fidelity numerical simulation; Lower left panel: the low-fidelity level data based on dimension reduction; Right panel: the prediction of quantity of interest obtained from numerical simulation results and low-dimensional data through Gaussian process regression. This study obtained a high-fidelity approximation from numerical simulation, and low-dimensional data is viewed as a low-fidelity approximation.

We have adopted a multi-fidelity Gaussian technique in the current work, which was first proposed by Kennedy and O'Hagan (see [10]) in a Bayesian environment. It can effectively

use cheap data by formulating some form of correction procedure to model the difference between cheap functions and expensive data. The main framework for solving such problems is Gaussian Processes (GPs) (see [18], [17], [11]). Through the work of Babaee *et al.* (see [2]), we can know that this is a non-parametric Bayesian machine learning technique. This supervised learning method provides a flexible prior distribution in the formula, is easy to handle, and can return a reliable posterior variance estimate. Besides, the Gaussian process belongs to a class of methods that utilize a kernel machine. These methods base on the auto-correlation kernel's parameterized study. Among them, the appropriate kernel hyperparameters are necessary, which will be calculated immediately based on the observed data, and will include deviations or modeling errors at each fidelity level. In other words, we can view that as a way to explore the interrelationship between high-fidelity and low-fidelity data. Although this method is not our original creation (see [6], [14]), we proved its practicality. Also, Our research shows that this method can significantly improve research efficiency.

1.3 Dimension reduction

On the other hand, low-fidelity data will be obtained by converting high-dimensional models into meaningful representations for dimensionality reduction. Dimension reduction is a very interesting research direction. Its mathematical principles have been completed in a very early period, but due to the limitations of experimental technology and computational performance at that time, the research in this area is mainly limited to mathematical derivation. Driven by the wave of the data/artificial intelligence (machine learning) era, it has once again become a point of interest.

Dimensionality reduction has the ability to discover internal structure from data. Snapshots (measurements) of many systems often show low-dimensional phenomena, so most of the variance/energy is contained in several modes calculated based on singular value decomposition (SVD). Ideally, the dimensionality of the reduced-order model should be close to the minimum quantity of parameters required to explain the system dynamics. Here, SVD

algorithm is one of the most important dimensionality reduction techniques for studying complex space-time systems. Such systems are usually represented by a series of equations that regulate the evolution of the quantity of targets in physical, engineering, and biological systems in time and space. SVD decomposes the original data (space-time) into hierarchically ordered modes in space and the time evolution coefficients (sequences) corresponding to each mode. In the context of fluid mechanics these modes are referred to as proper orthogonal decomposition (POD) modes. Each mode is orthogonal to each other. The order of the modes is sorted according to the magnitude of their captured energy (from high to low). These energies of these modes are represented by their respective eigenvalues; the expansion under this group of modes optimally captures the energy of the data: the r-truncated modal expansion is the optimal rank-r modal decomposition of the full-dimensional data. Importantly, the low-precision model produced by dimension reduction can significantly increase the calculation speed. This allows us to obtain the required low-precision data easily. Here, dimension reduction has two main goals: a) the ability to capture principal dynamic information of physical phenomena, and b) provide effective data compression methods for numerical simulation data sets. In fact, due to the wide application of dimension reduction, the function of dimension reduction has been sought after by people and has become a very important research topic.

Here is how we organize the rest of this thesis: In 2, we make a demonstration of the fluid mixing elbow problem. In 3, we make an illustration of the multi-fidelity framework and other technical details. In 4, we show the multi-fidelity results, and we made a conclusion in 5 with a summary.

2.0 Problem description and setup

Our data types are divided into two types; one is high-precision data directly obtained through simulation and low-precision data predicted through dimensional reduction method and least square regression based on a small amount of simulation sample data. In this chapter, we mainly introduce our basic physical model settings and the specific settings of numerical simulation.

First, we will demonstrate the methodology used to study the influence of boundary conditions on the flow. Thermal transfer studies aim to predict the temperature accurately and heat flux distribution in space, time, and boundaries, which occurs in various industrial applications such as aerospace, automotive, nuclear, oil, and gas applications. The temperature field is essential when considering thermal and material properties, which are the critical elements for the optimal design of thermal structures or fluid dynamics. In this study, we focus on flow with coupled heat transfer in an elbow in which two steams of flow with different temperatures are mixed. The temperature boundary at the hot inlet is variable. This variability is parameterized, and it could be due to uncertainty or other operating conditions. The outlet temperature is the primary quantity of interest in this study. The elbow setup schematic for the mixing of two flows are shown in Fig.2 (iii) and Mesh was built by ANSYS workbench, Fig 3 (i) shows the computational grid and Fig 3 (ii) shows a close-up view of the grid near the elbow joint.

As it is shown in Fig.2 (iii), a cold fluid flows into the pipe through inlet B and mixes with a hot fluid that enters through a smaller inlet (Inlet A) located at the elbow. After mixing at the elbow, the mixture flows out from outlet C. The prediction of temperature distribution on the surface of outlet C is the main purpose of our research. The boundary condition of inlet A and B are listed below. Here, T, u, and v respectively denote temperature, horizontal velocity, and vertical velocity.

$$\text{Inlet A: } u = u_a; v = v_a; T = T_a; \tag{2-1}$$

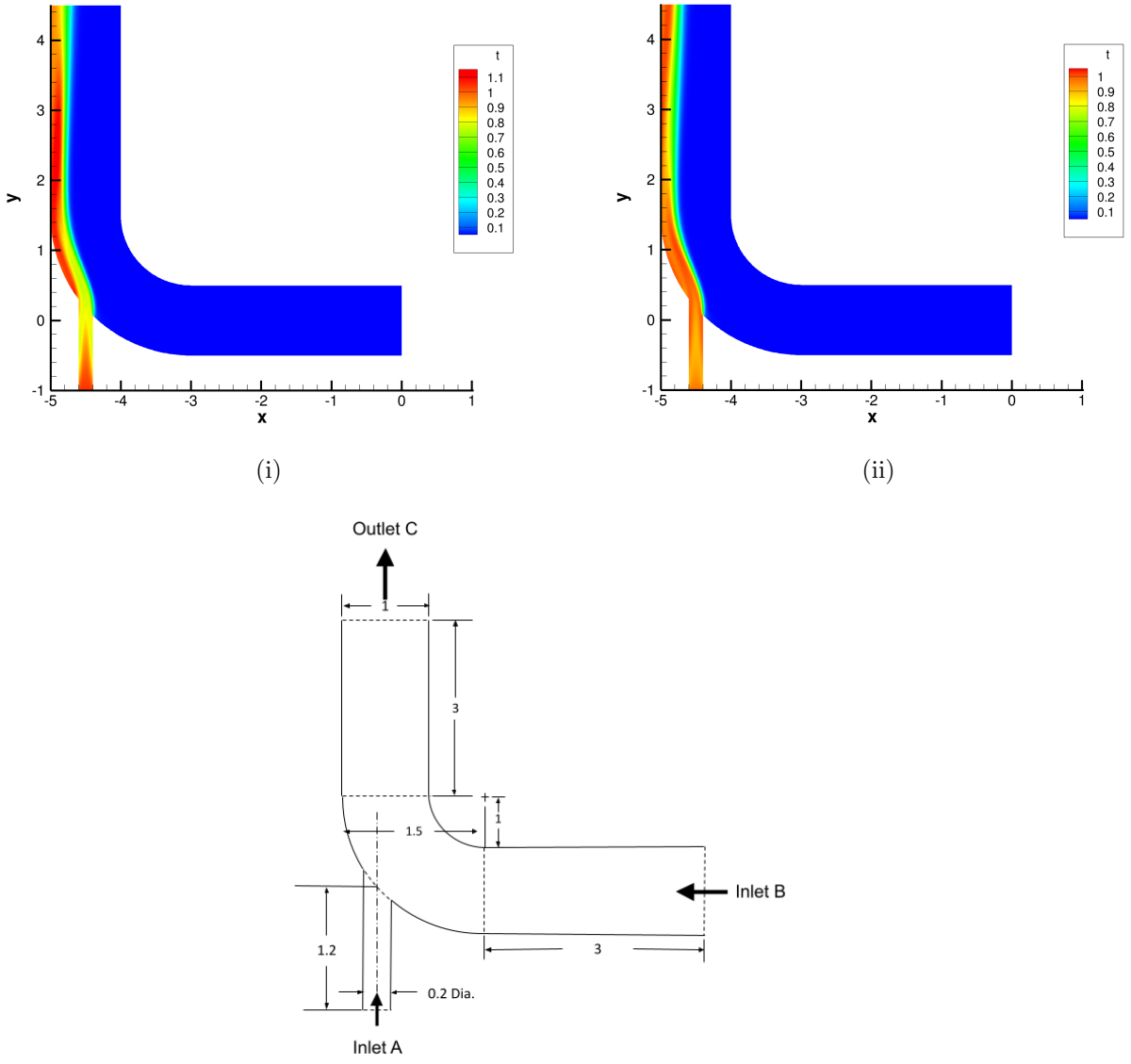
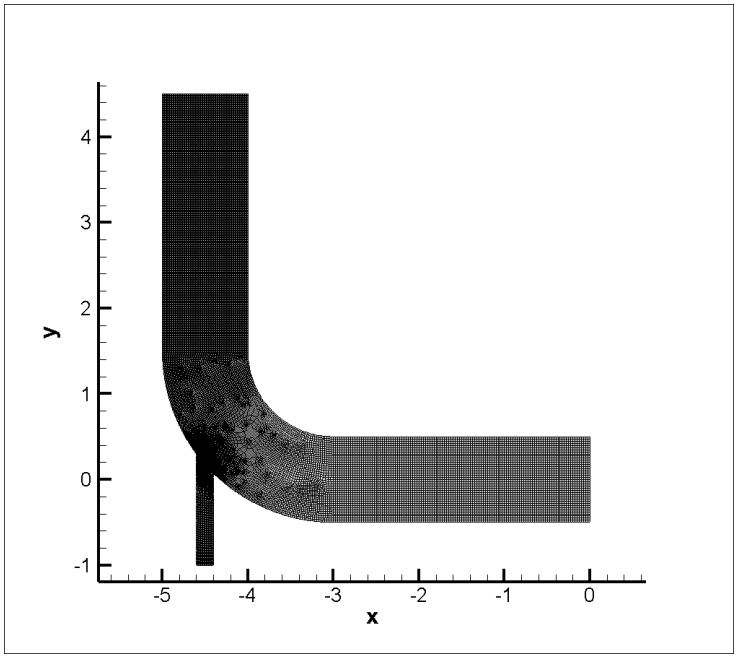
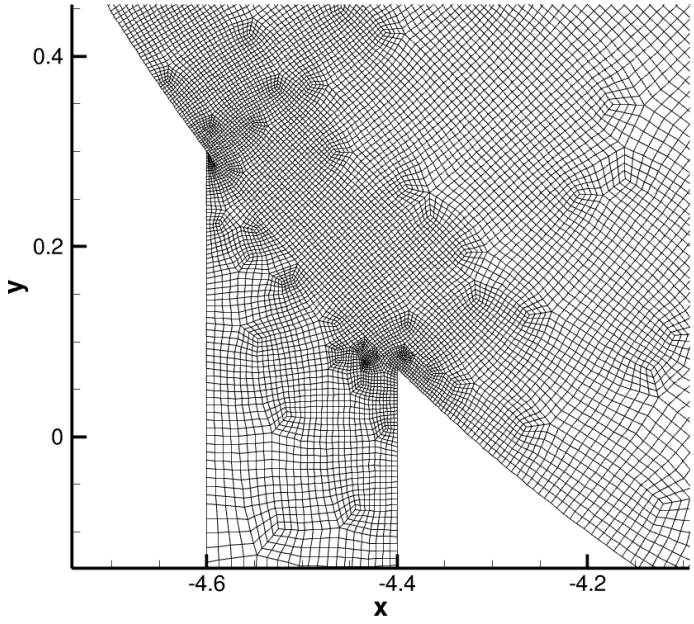


Figure 2. (i) Snapshots of temperature contours inside the elbow (time unit = 70). (ii) Snapshots of temperature contours insider the elbow (time unit = 72). (iii) Schematic of mixing elbow. From Fig (i) and (ii), the high-temperature fluid from the small inlet A is injected into the low-temperature fluid in the main pipe and mixed in the pipe. Heat transfer occurs at the elbow and the second half of the pipe.



(i)



(ii)

Figure 3. Unstructured quadrilateral grid; (i) two-dimensional view of elbow mixture. (ii) x-y view of the grid in the nearness of pipeline elbow grid. The grid become denser at the junction of the vertical pipeline and the main pipeline.

$$\text{Inlet B: } u = u_b, v = v_b, T = T_b; \tag{2-2}$$

Temperature of inlet A is specified by a time dependent parametric function: $T_a = T_a(\xi, \eta, t)$. Here, t represents time; $\{\xi, \eta\}$ are two parametric vectors: $\xi = (\xi_1, \xi_2, \xi_3 ... \xi_d)$,$\eta = (\eta_1, \eta_2, \eta_3 ... \eta_d)$ and d is the dimension of this vector. The initial condition for all cases considered here is identical and it is obtained by preforming a high-order numerical simulation simulation for a given choice of $\{\xi, \eta\}$ and steady state temperature profile at the inlet and the initial condition of:

$$u = 0, \quad v = 0, \quad T = 0; \tag{2-3}$$

We solve this case for a large quantity of time steps until we get a steady state result. Here, we use 1000 time steps to get the situation we want.Then, we will set that result as the initial condition of our problem. The objective of this study is to find the parametric dependence of the temperature as a function of space and time:

$$T = f(x, t, \xi, \eta). \tag{2-4}$$

Then based on the initial condition we got, we start with a pedagogical example involving the boundary condition equation:

$$T_a(t, \xi, \eta) = 1 + \sigma \sum_{i=1}^{n} [\frac{\xi_i}{i} \sin(it) + \frac{\eta_i}{i} \cos(it)]; \tag{2-5}$$

where ξ and η are independent normally distributed parameters and σ is a constant, which is assumed to be 0.1. Additionally, the high-order numerical simulation simulation training data $\{x_H, y_H\}$ are generated using spectral element solver $Nektar$.

In this example, we assign values for each parameter of inlet A and inlet B:

$$u_a = 0; \tag{2-6}$$

$$v_a = 1.5(1 - 100(x + 4.5)^2) \tag{2-7}$$

$$T_a = T_a(t, \xi, \eta); \tag{2-8}$$

$$u_b = -1 + 4y^2 \tag{2-9}$$

$$v_b = 0 \tag{2-10}$$

$$T_b = 0 \tag{2-11}$$

Here, the goal of all the concepts discussed in the following sections is to find an accurate expression of the equation 2-4 according to combining a small quantity of high-fidelity numerical simulation data with a large quantity of low-fidelity data. This method can serve as an excellent paradigm in predicting transient heat transfer.

3.0 Methodology

Computing the temperature for any parameter (ξ, η) is computationally prohibitive. To this end, we introduce a new method for efficient computation of temperature versus (x, t, ξ, η). As shown in Fig.3.3, our method is divided into three steps: (1) Dimension reduction based on SVD; (2) Least square regression; (3) multi-fidelity modeling. Here, there are two types of data types with different fidelity: (1) High-fidelity, which is obtained by performing numerical simulation; (2) Low fidelitywhich is obtained by performing dimension reduction. Additionally, we introduced how to combine the predictions of these two models to produce a multi-fidelity random response surface of temperature (see Fig. 4). Below, we explain how data from each of these sources is computed.

3.1 High-fidelity Model

First, we illustrate the mathematical model that governs the flow physics of the mixture elbow and the conservation of mass, momentum and energy in the mixture elbow are illustrated respectively by incompressible Navier-Stokes equations:

$$\nabla \cdot \boldsymbol{u} = 0, \tag{3-1}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \frac{1}{Re}\nabla^2 \boldsymbol{u}, \tag{3-2}$$

$$\frac{\partial T}{\partial t} + (\boldsymbol{u} \cdot \nabla)T = \frac{1}{RePr}\nabla^2 T, \tag{3-3}$$

In the above equations 3-1, 3-2 and 3-3, $\boldsymbol{u}$ represent the velocity of the fluids, which includes velocities in horizontal or vertical directions: u and v; pressure and temperature are respectively donated by p and T. Additionally, there are two non-dimensional parameters: Pr and Re: (1) Reynolds quantity: $Re = u_{ref} l_{ref}/\nu$ where ν represents kinematic viscosity. Here,

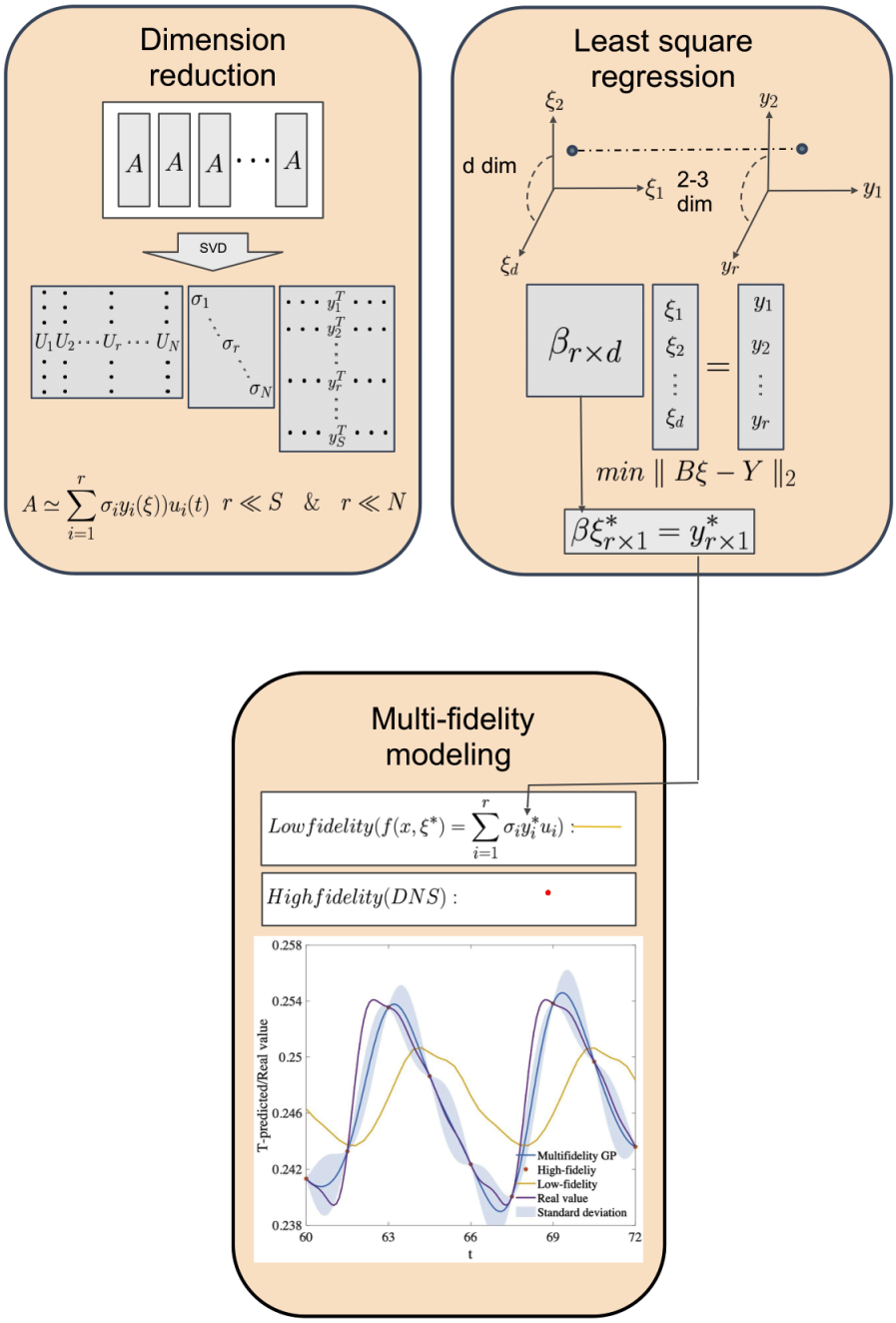


Figure 4. Based on the existing numerical simulation data, our approach is split into three steps: 1)Upper left panel: Decompose and reduce the dimension of the existing numerical simulation data set A. 2)Upper right panel: Establish a low-fidelity model for a particular ξ^* by predicting the right-singular vector of the corresponding data $f(x, \xi^*)$. 3) Lower panel: For a given variable xi^*, combine its corresponding numerical simulation data f_{HF} and low-precision model f_{LF} to generate a multi-precision framework for flow temperature.

u_{ref} can be substituted by the average velocity of inlet surface A, u_a or the average velocity of inlet surface B, u_b and the choices of l_{ref} are the inlet diameter D_a and D_b. (2) Prandtl quantity: $Pr = \nu/\alpha$ where α is the thermal diffusivity.

Numerical Simulation are performed to solve equations 3-1, 3-2 and 3-3. The boundary conditions assigned to the first and second fluid flows enforced at the left and right inlets are demonstrated by equations 2-1 and 2-2. We solve the above equation with a relatively low Reynolds quantity and a two-dimensional domain, as shown in Fig.2 (iii). We obtain our high-fidelity approximation data using of spectral/hp element method implemented in Nektar, in which we use spectral polynomials in each element, and time splitting method to advance the equations in time. The spectral/hp method is a methodical solution with the geometric flexibility of the classical finite element method and the exactness of continuum elements. For more detailed instructions, see [9]. The spectral/hp element method has been used in a wide range of applications and more relevantly in highly turbulent forced convection film cooling problem and mixed thermal convection problem [2].

In order to obtain the initial conditions of the flows in the elbow, we assign $\{\xi, \eta\}$ equal to $\{\xi_0, \eta_0\}$ and define the initial condition of velocity and temperature u_0, v_0 and T_0. We run this initial condition simulation with a given Pr and a specific spectral polynomial order. Additionally, we set enough time steps with time intervals of Δt so that the result showed that the flow in the elbow had already become a stable state. Hence, the final snapshot file is treated as the initial temperature condition of subsequent simulation.

3.2 Low-fidelity Model

The high dimensional dependence of temperature to parametric variations means that this problem suffers from the curse of dimensionality if we use traditional discretizations schemes in the parametric space, in which case the computational complexity increases exponentially with the dimension. We describe a method for building a low fidelity model to reduce the size of the matrix. The goal is to find a ROM for this data set with reasonable accuracy. The aim of ROM is to discover the intrinsic structure from data. The goal of

dimension reduction is to map high-dimensional data into a meaningful representation of reduced dimension. This method is achieved by combining Singular value decomposition (SVD) with the least-squares regression.

3.2.1 Singular Value Decomposition

SVD is a method of dimension reduction based on the principle of least squares technology (see [7]). Snapshots (measurements) of many nonlinear dynamical system often exhibit low-dimensional phenomena, so that the majority of variance/energy is contained in a few modes computed from SVD. It projects the feature matrix A to a low-dimensional space to form a low-dimensional matrix A, which means it minimizes least square of the difference between the raw data A and the low-dimensional representation of A. In this study, we are interested in the outflow temperature $f(x, \xi)$ where ξ is viewed as S set of random variables, so the matrix is constructed as:

$$A = \begin{pmatrix} f(x^1, \xi_1) & f(x^1, \xi_2) & \cdots & f(x^1, \xi_S) \\ f(x^2, \xi_1) & f(x^2, \xi_2) & \cdots & f(x^2, \xi_S) \\ \ddots & \vdots & \ddots & \vdots \\ f(x^N, \xi_1) & f(x^N, \xi_2) & \cdots & f(x^N, \xi_S) \end{pmatrix}, \tag{3-4}$$

Where x can be a multiple dimensional vector, presenting the space points and time step t. For example, we assume $x = \{x_1, x_2\}$. Here x_1 present the position on X axis in Fig.2 (iii) and x_2 present the time step, thus the matrix A can be construct in this form:

$$A = \begin{pmatrix} f(x_1^1, x_2^1, \xi_1) & f(x_1^1, x_2^1, \xi_2) & \cdots & f(x_1^1, x_2^1, \xi_S) \\ f(x_1^2, x_2^1, \xi_1) & f(x_1^2, x_2^1, \xi_2) & \cdots & f(x_1^2, x_2^1, \xi_S) \\ f(x_1^3, x_2^1, \xi_1) & f(x_1^3, x_2^1, \xi_2) & \cdots & f(x_3^1, x_2^1, \xi_S) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_1^{N_1}, x_2^1, \xi_1) & f(x_1^{N_1}, x_2^1, \xi_2) & \cdots & f(x_1^{N_1}, x_2^1, \xi_S) \\ f(x_1^1, x_2^2, \xi_1) & f(x_1^1, x_2^2, \xi_2) & \cdots & f(x_1^1, x_2^2, \xi_S) \\ f(x_1^2, x_2^2, \xi_1) & f(x_1^2, x_2^2, \xi_2) & \cdots & f(x_1^2, x_2^2, \xi_S) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_1^{N_1}, x_2^{N_2}, \xi_1) & f(x_1^{N_1}, x_2^{N_2}, \xi_2) & \cdots & f(x_1^{N_1}, x_2^{N_2}, \xi_S) \end{pmatrix}, \tag{3-5}$$

The SVD of matrix A results in:

$$A_{N\times S} = U_{N\times N}\Sigma_{N\times S}V^T_{S\times S}, \tag{3-6}$$

where

$$U_{N\times N} = \begin{pmatrix} u_1(x^1) & u_2(x^1) & \cdots & u_N(x^1) \\ u_1(x^2) & u_2(x^2) & \cdots & u_N(x^2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(x^N) & u_2(x^N) & \cdots & u_N(x^N) \end{pmatrix}, \tag{3-7}$$

$$V_{S\times S} = \begin{pmatrix} v_1(\xi_1) & v_2(\xi_1) & \cdots & v_S(\xi_1) \\ v_1(\xi_2) & v_2(\xi_2) & \cdots & v_S(\xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ v_1(\xi_S) & v_2(\xi_S) & \cdots & v_S(\xi_S) \end{pmatrix}, \tag{3-8}$$

where N_1 is the quantity of x_1 and N_2 is the quantity of x_2 and S represents the sample quantity with different sets of random variable ξ and $S < N$. The matrices U and V are two orthogonal matrices, where $U^T \times U = I$ and $V^T \times V = I$. The columns in matrix U are called the left singular vectors, while the matrix V columns are the right singular vectors. Among them, $u(x)$ and $v(\xi)$ respectively correspond to the variables x and ξ we sample. The matrix Σ is a diagonal matrix, where $\Sigma_{i,j} = 0$ for $i \neq j$, and the diagonal elements of Sigma are the singular values of matrix $A_{N\times S}$. The singular values in the matrix Σ are sorted by descending order such that $(\sigma_{1,1} > \sigma_{2,2} > ....\sigma_{S,S})$. For such a case, the SVD basis is typically truncated at a predetermined cut-off, such as when the model basis contain 90% of the variance, so that only the first r-modes(r-rank truncation) are kept. Thus it is possible to remove the singular values less than $\Sigma_{r,r}$, where $r \ll S$ and $r \ll N$. This way, the three matrices are reduced to:

$$A_{N\times S} \simeq U'_{N\times r}\Sigma'_{r\times r}V'^T_{S\times r} \tag{3-9}$$

Matrix $A_{N\times S}$ is expected to be close enough to matrix $A_{N\times S}$ that satisfies $||A - \tilde{A}||_2 < \sigma$. Matrix $U'_{N\times r}$ is a reduced matrix of $U_{N\times S}$, $\Sigma'_{r\times r}$ is the reduced version of $\Sigma_{S\times S}$ from N to r dimensions and $V'_{S\times r}$ is the reduced matrix of $V_{S\times S}$. Here, we can use $A'_{N\times S}$ to present the low order surrogate of the origin measurement. So, for a new set of variable ξ^*, we can build its corresponding low-fidelity measurements section A_{LF} by predicting the corresponding

right singular vector $v(\xi^*)$.We will talk about the whole process in 3.2.4. In the next step, we hope to be able to predict the corresponding low-fidelity measurements of $v(\xi^*)$ for the given new variable ξ^* through least squares regression.

3.2.2 The Least Square Regression

As the most widespread and mature method of machine learning out there, linear regression is a linear approach for assuming a relationship between the dependent variable y and one or more independent variables ξ. Here, y represents the element in matrix Y columns, right singular vectors of origin matrix A, mentioned in the previous section. Therefore, for a new set of variables ξ, we can predict the accurate expression of its corresponding y via a fitted regression line:

$$y = \beta_0 + \beta_1\xi_1 + \beta_2\xi_2 + \cdots + \beta_d\xi_d + \epsilon \tag{3-10}$$

In equation 3-10, $\{\beta_1, \beta_2, ...\beta_d\}$ is the slope of the line, β_0 is y-intercept. Hence, β reflected the relationship between y and ξ. However, for any specific observation, the actual value of $\tilde{y}$ can deviate from the predicted value.The deviations between the actual and predicted valued are treated as errors ϵ. The better the line fits the data, the smaller the errors. Thus, in this section, we want to determine the value of β so we can predict y for any new value of variable ξ.

Here, we use S sets of data points $\{\xi_{d\times 1} to, y\}$ to find the corresponding ξ. we square all errors for each point and find a line that minimizes this sum of the squared errors. This method helps us find value of the intercept and slope coefficient that minimize the sum of the squared errors.

$$\sum_{j=1}^{S} \epsilon_j^2 = \sum_{j=1}^{S} (Y_j - \overline{Y}_j)^2 \tag{3-11}$$

Additionally, as shown in equation 3-8, the dimension of V' is larger than 1. So, For illustrating the derivation process clearly, we use matrix $y_{r\times S}$ represent ${V'_{S\times r}}^T$ then we can derive:

$$A_{N\times S} \simeq U'_{N\times r}\Sigma'_{r\times r}Y_{r\times S}$$

.

First, considering β is y-intercept, we define the first row of matrix equal to 1 and we donate constant term by β_0 so that we can make sure every item in the formula is considered.

$$y = \begin{pmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_S \end{pmatrix}_{r\times S}, \quad X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \xi_{11} & \xi_{12} & \cdots & \xi_{1S} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{d1} & \xi_{dS} & \cdots & \xi_{dS} \end{pmatrix}_{(d+1)\times S}, \tag{3-12}$$

and

$$\beta = \begin{pmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \cdots & \hat{\beta}_d \end{pmatrix}_{r\times(d+1)}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_r \end{pmatrix}_{r\times S}, \tag{3-13}$$

In the first index in the matrix $X_{(d+1)\times S}$, d refers to the dimension of variable ξ (in columns) and the second index S refer to the feature quantity(in row). Thus, we can rewrite equation 3-10 as:

$$y = \beta X + \epsilon. \tag{3-14}$$

where β is a $\{r \times (d+1)\}$ vector we want to estimate and ϵ is a $n \times 1$ vector of overlooked errors. Then, ϵ can be written by

$$\epsilon = \beta X - y \tag{3-15}$$

and

$$E(\beta) = \sum_{j=1}^{S} \epsilon_j^2 = \sum_{j=1}^{S} \epsilon_j' \epsilon_j = \sum_{j=1}^{S} (y_j - \beta X_j)^T (y_j - \beta X_j) \tag{3-16}$$

Since the least squares estimator is obtained by minimizing $E(\beta)$, we can find β by the following differential equation:

$$\frac{\partial E}{\partial \beta} = -2yX^T + 2\beta X X^T. \tag{3-17}$$

Setting these derivatives equal to zero, we obtain:

$$\beta X X^T = yX^T \tag{3-18}$$

Therefor, β is obtained as follows:

$$\beta = yX^T(XX^T)^{-1} \tag{3-19}$$

This equation require the inverse of (XX^T) exists, which means that $d \leq S$. For $\xi_{(d\times S}$, the dimension of variable ξ must be smaller than the quantity of variable ξ. Thus, when we have a new set of variable ξ^*, we can get the corresponding $y^*(\xi^*)_{1\times r}$.

$$y^* = \beta\xi^* \tag{3-20}$$

Here, y^* represent the estimation of the vector $v(\xi^*)$. Finally, according to equation 3-9, we can obtain the outflow temperature $f(x, \xi^*)$ by:

$$f(x, \xi^*) = \sum_{i=1}^{r} \sigma_i y_i u_r(x) \tag{3-21}$$

3.2.3 Cross-Validation Test

Here, a cross-validation test is implemented for least square regression models in this section. We view matrix y and ξ as our raw data sets:

$$y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1S} \\ \vdots & \vdots & \ddots & \vdots \\ y_{r1} & y_{r2} & \cdots & y_{rS} \end{pmatrix}_{r\times S} \tag{3-22}$$

and split it, consisting of S columns of vector, to two separate data sets: the first one is actually used for training the least square regression model with $S_t = a \times S; a < 1$ points randomly chosen from raw data set $\{\xi, y\}$ and the other one which consist of $S_{cv} = S - S_t$ remaining data sets $\{\xi^{cv}, y^{cv}\}$ that are used to validate the model. The error is the L_2 error defined as:

$$\epsilon^2 = \frac{1}{N_{cv}} \sum_{i=1}^{N_{cv}} (\beta\xi_i^{cv} - y_i^{cv})^2, \tag{3-23}$$

where matrix β with the size of $r \times d$ is estimated for equation 3-19 that base on the first data set $\{\xi^{cv}, y^{cv}\}$ and y_{cv} are obtained from singular value decomposition and consist of S_{cv}

columns of remaining vectors. The training process has been described in previous section. The S_t columns of training vectors are selected randomly out of $y_{r\times S}$. In order to estimate the error more accurately, we will repeatedly build the training data set, and the cross-validation set as described above and calculate the overall error of the model in the following way. Firstly, we build M training sets and every one randomly obtained from the raw data sets $\{\xi, y\}$. Among M training sets, each training sets contain S_t sets of points selected randomly from ξ, Y. Meanwhile, we have M cross-validation sets consist of S_{cv} remaining data points, corresponding to each training sets. Each model's error was calculated based to the equation 3-23. Then, we can calculate the mean and variance of the error estimate for the array of M models, signified by $\overline{\varepsilon^2}$ and σ_{ε^2} respectively. Finally, We rerun the above process for different value of S_t, which is the quantity of training sets.

3.2.4 Proposed Approach

In this section, we demonstrate how we deploy the SVD technique to construct lower-dimensional space for different feature for the low fidelity data part of multi-fidelity models. We performed S sets of training simulations in Nektar, based on the Spectral/hp element method, to generate a sufficient quantity of snapshots. The only factor that influences the differences in the results of these sets of simulation is inlet A temperature, $T_b(t, \xi)$. According to MATLAB function 'randn', which return a random scalar drawn from the standard normal distribution, we can got S sets of sample with different $\xi = \{\xi_1, \cdots, \xi_s\}$.

In the next chapter, we will discuss two cases to test the effectiveness of our method in one-dimensional and two-dimensional problems. In the first case, our objective function is the change in the average temperature of the outlet surface over time$f(\xi, x_1)$. In the second case, our attention is shifted to the entire outlet surface temperature instead of the average temperature, so the objective function is: $f(\xi, x_1, x_2)$.

Now, we start to build our matrix A whose elements correspond to the temperature on outflow surface. Additionally, one column in matrix A presents all time steps in one sample and each column in matrix A corresponds each set of sample with certain ξ in all time steps. Specifying the temperature snapshot as $f_1(t), f_2(t)$, etc., the $N \times S$ temperature snapshots,

matrix A, is constructed as:

$$A = [f_1(t), \quad f_2(t), \quad ... \quad f_S(t)]. \tag{3-24}$$

Thence, A is a matrix with N rows and S columns, respectively corresponding to the quantity of snapshots and the quantity of objects. Then, we performed singular value decomposition(SVD) of this matrix, the left singular vector U whose column is corresponding the N time step. Also, the right singular vector V is corresponding to the S samples with different ξ value.

$$A = \begin{bmatrix} f_1(t)^1 & f_2(t)^1 & \cdots & f_S(t)^1 \\ f_1(t)^2 & f_2(t)^2 & \cdots & f_S(t)^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_1(t)^N & f_2(t)^N & \cdots & f_S(t)^N \end{bmatrix} = U_{N\times N}\Sigma_{N\times S}V^T_{S\times S} \tag{3-25}$$

Since only ten samples can not give us enough information, we need a considerable amount of low-fidelity samples. So, we denote:

$$Ur = [u_1, u_2, \cdots u_r], r = diag(\sigma_1, \sigma_2, \cdots, \sigma_r), Vr = [v_1, v_2, \cdots, v_r] \tag{3-26}$$

Then, we can get:

$$A \simeq U_r\Sigma_r V_r^T = \sum_{i=1}^{r} u_i\sigma_i v_i^T \tag{3-27}$$

Here, the value of r determines how many percent of the accuracy of the matrix A can be captured. Then, We use the least square method to predict the new right singular vector V for the new ξ. According to Linear Regression, we can build the equation with vector $\xi = \{\xi\}_i^d$:

$$y = \beta_{r\times d}\xi_{d\times S} \tag{3-28}$$

Here, y corresponds to the estimation of the right singular matrix of A and we view ξ, V_r^T as the training data points to estimate the value of β. With the help of least squares regression, we can calculate matrix $\beta_{r\times d}$ that can reveal the relationship between ξ and V^T. Then, we can do a prediction when we assign a new set of values to ξ^*. We set

$\xi^* = \{\xi_1^*, \cdots, \xi_d^*\}$; After entering the above formula 3-28, we can obtain the new coefficient f_{LF}^* at each snapshots.Then, we can predict temperature of elbow mixture by

$$\overline{f_{LF}^*} = U_r \Sigma_r y^* \tag{3-29}$$

We can quickly obtain a large amount of cheap and effective low-fidelity data sets to assist us in the next step of multi-precision modeling through the above method.

3.3 Multi-Fidelity Modeling

This section demonstrates the Gaussian Process (GP) regression method with kernels for vector-valued functions, which depend on any quantity of variables. GP is a probability measure of a function, so that the function values on any set of input points have a joint Gaussian distribution (see [3]). That is to say, for any group of observations with a joint Gaussian distribution, the distribution of the subset conditioned on the other conditions is also Gaussian. Based on this characteristic and the above facts, we can make predictions at unknown points through previous observations. The GP used here is the basic component of the Bayesian method. In other words, GP regression can be interpreted by Bayes' rule as a posterior state that represents a person's understanding of unknown functions, and the posterior state depends on the observed data. In addition, since the posterior is also GP, the expected value and covariance of any new point set can be obtained through analysis. In general, the Bayesian non-parametric nature of GP, the analysis efficiency of its processing, and the natural expansion of multiple fidelity settings based on this work led to this choice.

Inspired by [16] and [11], the recursive processes of multi-fidelity Gaussian regression methods(GPR) is discussed. This approach can serve as a basis for applying to multi-objective problems in the proposed approach, where the amount of interest might be a multi-dimensional vector or even a continuous function. As shown in Fig.5, we consider two fidelity levels of data. On one hand of the assortment of fidelity, we have high-fidelity models that we place complete trust in, which is expensive to evaluate. For instance, it could be collected from numerical simulation, experiments, etc. However, we may face limited

availability of high-fidelity data. On the other side, we may also have access to cheaper surrogate models, which are fast to compute but less trustworthy(e.g., solvers of potential flow, analytical formulations, etc.). Exploring the connection between the two in several instances helps one create an accurate representation of the response surface effectively by carrying out relatively few evaluations of an expensive high-fidelity model and more evaluations of a cheaper substitute. Actually, the formulation can be extended to $n > 2$ levels of fidelity in a straightforward manner(see e.g., [6]) and to nonlinear auto-regressive schemes using deep GP (see e.g., [14]).In the framework of multivariate Gaussian regression, a binding mechanism that undertakes information fusion task naturally arises. Next, in the sense of traditional co-kriging, we formulate a general methodology that can simultaneously address multi-fidelity in physical models as well as multi-fidelity in probability space.

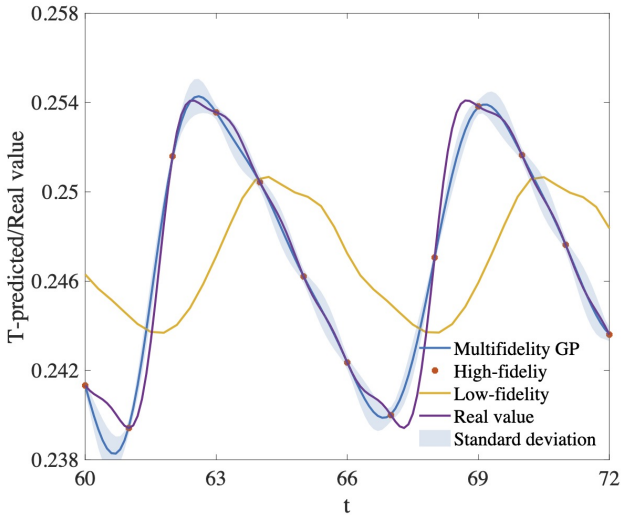


Figure 5. In present study, the linear regression model based on dimensional reduction is regarded as the heat transfer problem's low-accuracy estimation. At the same time, the high-fidelity observation data is driven by numerical simulation tool. The uncertainty that is specified by blue area shows two times standard deviation ($\pm 2\sigma[T]$) around the estimation work for the multi-fidelity models.

3.3.1 Multi-Fidelity Scheme

Co-Kriging is a more efficient way of estimating the correction processes. The following description is based on the process that has been explained in [15]. Co-Kriging can formulate a particular form of correction process which can couple a greater quantity of cheap data with a small amount of expensive data to enhance the accuracy of a surrogate of the expensive function. with two sets of data, we begin the co-Kriging formulation by concatenating the sample positions to give the combined set of sample points.

$$x = \begin{pmatrix} x_L \\ x_H \end{pmatrix}, = \begin{pmatrix} x_L^{(1)} \\ \vdots \\ x_L^{(n_L)} \\ x_H^{(1)} \\ \vdots \\ x_H^{(n_H)} \end{pmatrix}, \tag{3-30}$$

where, x_L and x_H are respectively corresponding to the predictor variable of the low- model and the high-fidelity model and represents the D-dimensional input space. For temporal multi-fidelity models, D = 1, where x represents time. As a realization of a Gaussian random field $f_L(x)$, the key concept here is to model scattered observations of a field $f_H(x)$. Here,$f_H(x)$ and $f_L(x)$ can be respectively viewed as the low fidelity level and high-fidelity level measurements of outflow temperature $f(x, \xi)$. We approximate the expensive data using the auto-regressive model by multiply.ing the inexpensive data by a constant scaling factor and adding them with Gaussian operation $\varepsilon(x)$, which represents the difference between $f_H(x)$ and $f_L(x)$.

$$f_L(x) = u_L(x) + \varepsilon_L \quad and \quad f_H(x) = \rho u_L(x) + \delta(x) + \varepsilon_H; \tag{3-31}$$

where $u_L(x)$ and $\delta(x)$ are assumed to be two independent random functions, each one represented by a GP in the form of

$$u_L(x) \sim \mathcal{GP}(0, k(x, x'; \theta_L)) \quad and \quad \delta_H(x) \sim \mathcal{GP}(0, k(x, x'; \theta_\delta)); \tag{3-32}$$

and the unprejudiced errors in the low-fidelity and high-fidelity data are respectively donated by ε_L and ε_H. As pointed out by H.Babaee[2], through the cross-correlation parameter ρ in 3-31,the contribution of the low-fidelity data set into the high-fidelity estimation is captured. The model error or measurement noise defined by a Gaussian process, $\varepsilon(x)$ might corrupt these observations,Which leads to the observation model of the equation 3-32. Additionally, the multi-fidelity strategy given by equation 3-31 brook and gain an understanding of the System deviation in low fidelity measurement. This bias is modeled by the scaling factor $1/\rho$ and the bias correction of $\delta(x,t)/\rho$. In cases where the high-fidelity data have systematic bias, the uncertainty has to be modeled explicitly. In most cases presented in this paper, We use the squared exponential kernel (for posterior predictions) as specified by

$$k(x,x',\theta) = \sigma^2 exp(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_i - x_i')^2}{\theta_i^2}) \tag{3-33}$$

where we view δ and θ_i as hyper-parameters; therefore, the kernel at each fidelity level contains D + 1 hyper-parameters. In addition to the hyper-parameters introduced by δ_{nL} δ_{nH} and ρ, a total of $2D+5$ hyper-parameters will be generated. Here, D present the column quantity of x and we can view D as the dimension of predictor variable x.

3.3.1.1 Training

we learn hyper-parameters from low-fidelity and high-fidelity observations. Here, the low-fidelity observations are denoted by $\{x_L,f_L\}$, where x_L is a $n_L \times D$ input matrix and f_L is a $n_L \times 1$, representing reduced order modeling data pointsand n_L is the quantity of low-accuracy data. Similarly, we denoted the high-precision data by $\{x_H, f_H\}$, where x_H with the size of $n_L \times D$ are high-fidelity input matrix and f_H with the size of $n_H \times 1$ are corresponding numerical simulation results, and n_H is the quantity of high fidelity data. The quantity of low-fidelity data points is much larger than the quantity of high fidelity data points, $n_L \gg n_H$.

Since $u_L(x)$ and $\delta(x)$ are independent GP, using the multi-fidelity scheme given by equation $3-31$, it is straightforward to observe that the low-fidelity and high-fidelity observations have a joint Gaussian distribution given by $y \sim \mathcal{N}(0, K)$, where

$$f = \begin{bmatrix} f_L \\ f_H \end{bmatrix} \quad and \quad K = \begin{bmatrix} k_{LL}(x_L, x_L) + \delta_{n_L}^2 I & k_{LH}(x_L, x_H) \\ k_{HL}(x_H, x_L) & k_{HH}(x_H, x_H) + \delta_{n_H}^2 I \end{bmatrix}, \tag{3-34}$$

where f is a $n \times 1$ vector that contains the low-fidelity and high-fidelity data at points $x = [x_L, x_H]$ with the size of $n \times D$, n is the quantity of high-fidelity and low-fidelity data points, that is, $n = n_H + n_L$, and

$$k_{LL}(x_L, x_L) = k(x_L, x_L; \theta_{_L}), \tag{3-35}$$

$$k_{LH}(x_L, x_H) = \rho k(x_L, x_H; \theta_L), \tag{3-36}$$

$$k_{HL}(x_H, x_L) = k_{LH}(x_L, x_H)^T, \tag{3-37}$$

$$k_{HH}(x_H, x_H) = \rho^2 k(x_H, x_H; \theta_L, \theta_\delta, \rho) + k(x_H, x_H; \theta_\delta), \tag{3-38}$$

where $k_{LL}(x_L, x_L)$ is an $n_L \times n_L$ matrix that reflects the relationship between the low-fidelity points, the relationship between the low-accuracy points and high-accuracy points is represented by $k_{LH}(x_L, x_H)$ with the size of $n_L \times n_H$, the correlations between high-fidelity points are represented by matrix $k_{HH}(x_H, x_L)$ with the size of $n_H n_H$. Here, what we should pay attention to is $k_{HL}(x_H, x_L) = k_{LH}^T(x_L, x_H)$. Finally, The negative log-likelihood $\mathcal{L}(y|\theta_L, \theta_\delta, \sigma_{n_L}, \sigma_{n_H})$ is expressed by

$$-log\mathcal{L}(f|\theta_L, \theta_\delta, \rho, \sigma_{n_L}, \sigma_{n_H}) = \frac{1}{2} f^T K^{-1} f + \frac{1}{2} log|K| + \frac{n_L + n_H}{2} log(2\pi) \tag{3-39}$$

The hyper-parameter $(\theta_L, \theta_\delta, \rho, \sigma_{n_L}, \sigma_{n_H})$ are derived by minimizing the above function, which means we maximize the likelihood. The trust-region method can solve such a minimizing problem. The size of kernel matrix K is $n \times n$ full matrix; Thus, its storage requirement is O(n2), and the complexity of its inversion operation is O(n3). However, this is prohibitive

that we used this approach for large data sets. Le Gratiet [12] has suggested a recursive auto-regressive scheme to address this issue. Here, the two fidelity data sources can be decoupled and reconstructed equivalently in two independent recursive Gaussian regression problems, which requires nested multi-fidelity data sets.

3.3.1.2 Prediction

Once we found the hyper-parameters, we can use the probabilistic model for making prediction of mean value and the associated uncertainty at new observed space points and times. These estimation can be calculated for any random point. Also, they are independent of the position of the training data. From this perspective, the predictions of the multi-fidelity model are continuous. Here, we treat the resulting posterior distribution at the high-fidelity level as the prediction of the multi-fidelity model. The joint Gaussian prior distribution on the low-fidelity/high-fidelity observations(y) can be conditioned, which correlate to the probabilistic prediction at random points x_*. For this reason, It can be observed that the joint probability density function(PDF) of the new prediction points and the training points is also a Gaussian distribution given by

$$\begin{bmatrix} u_H(x_*) \\ f \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{HH}(x_*, x_*) & a \\ a^T & K \end{bmatrix} \right) \tag{3-40}$$

where $a = g(x_*, x; \theta) = [K_{HL}(x_*, x_L), K_{HH}(x_*, x_H)]$ and x_L, x_H and X correspond to equation $3-30$. The above-mentioned relationship between the prediction points and multi-fidelity data sets can be utilized to explore the conditional distribution to make predictions:

$$p(u_H(x^*)|f) = \mathcal{N}(\overline{u}(x_*), V_f(x_*)) \tag{3-41}$$

where $\overline{u}(x_*) := aK^{-1}f$ is the posterior mean (i.e., predictions), and predictive variance is derived by $\sigma_H^2 := diag(k_{HH}(x_*, x_*) - aK^{-1}a^T)$, and $\pm 2\sigma_H$ reflect the uncertainty of the multi-fidelity predictions.

4.0 Results

4.1 Prediction Result With 1-D Multi-fidelity Framework

The study of the transient heat transfer inside of elbow mixture was carry out from $1-D$ multi-fidelity framework. Form this, we can understand the basic of the outlet surface temperature change with time and the effect of data sources with different fidelity level on the multi-precision Gaussian regression framework.

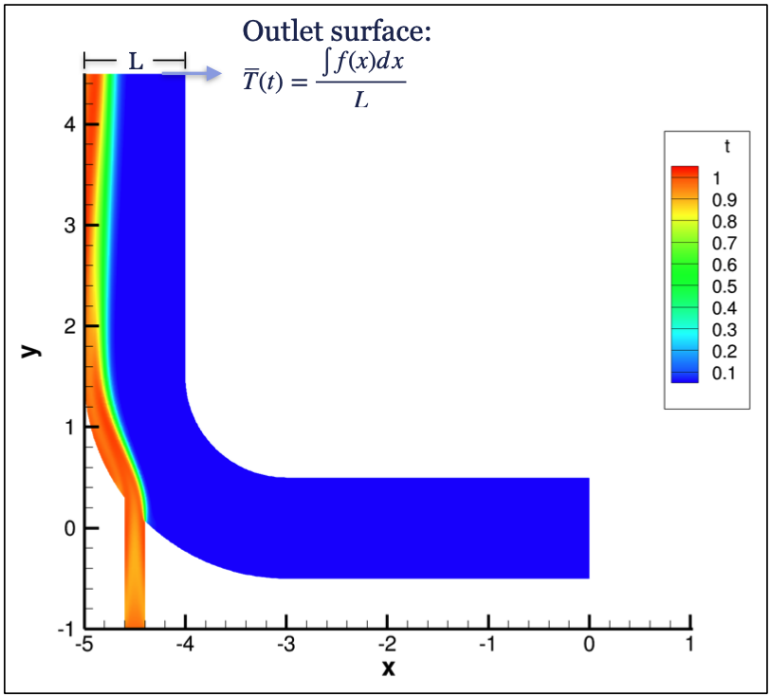


Figure 6. The study domain of multi-fidelity modeling. In $1-D$ multi-fidelity framework, the prediction target on the average temperature of outlet surface.

The computation domain for 1D multi-fidelity modeling is shown in Fig.4.1. In our data set $\{x, f\}$, f represent average temperature of outlet surface in different snapshots:

$$f = \overline{T}(t) = \frac{\int f(x_1)_t dx}{L} \tag{4-1}$$

where x_1 is space point on the outflow surface of elbow and $T = f(x_1)_t$ represents the outlet surface temperature with different time steps. We assume that the low- and high-fidelity training data $\{x_L, f_L\}$ are generated according to equation 4-4 and $\{x_H, f_H\}$ based on numerical simulation, Nektar simulation. In term of Nektar simulation, we start with the boundary condition of inlet A:

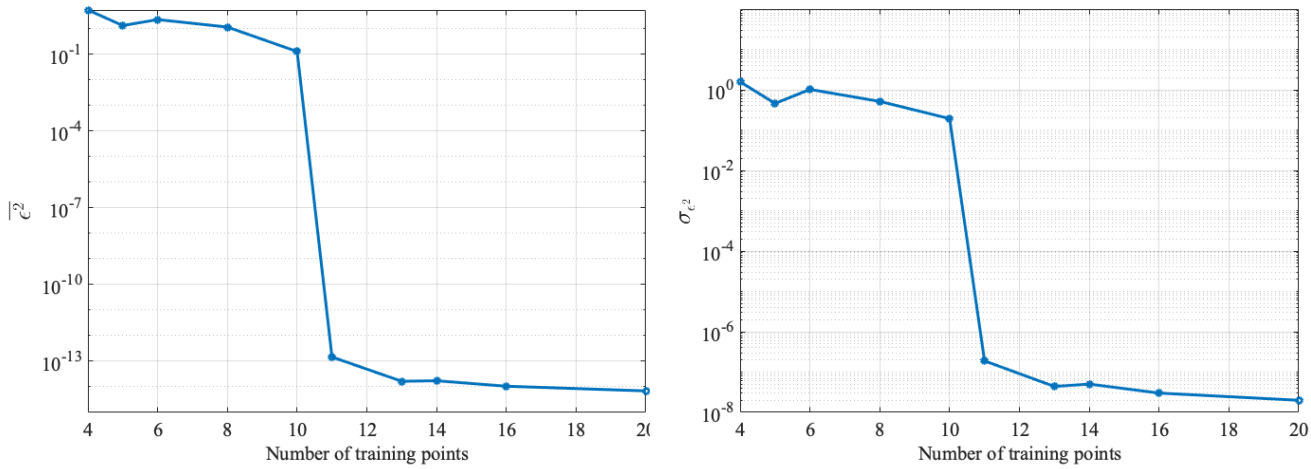
$$T_a = 1 + \sigma \sum_{i=1}^{n} [\frac{\xi_i}{i} sin(it) + \frac{\eta_i}{i} cos(it)]; \tag{4-2}$$

Here,we take $n = 5$ and n denote the dimension of ξ and η. In our study, we need to process multiple sets of training samples by Nektar with different settings of $\{\xi,\eta\}$ for building high- and low-fidelity data, respectively. On the one hand, We will take S settings of $\{\xi,\eta\}$ for building matrix A and the specific value of S will be in next section, on the other hand, we set $\{\xi, \eta\} = \{0.5, 0.5, \cdots .0.5\}_{1\times 10}$ as a particular case we will predict by multi-fidelity modeling. For this specific case, we will calculate the corresponding f_L and f_H for building multi-fidelity modeling. Here, we take $n_2 = 41$ and change the value of n_1 to see its influence on prediction result in later section. From Fig.(iii), we can get the characteristic length and velocity for inlet A and inlet B. The Reynold quantity for inlet A is 200 and the Reynold quantity for inlet B is 660. The Prandtl quantity is $Pr = 1.0$.

4.1.1 The Cross-validation For Least Square Regression

In this section, our goal here is to create a data set y whose column quantity, the quantity of sample, is sizeable enough to provide a quantification of predictive precision for the trained surrogate models, while conserving a rational computational forecast as each sample was derived from numerical simulation. Here, we use cross validation method to evaluate the accuracy of the least regression model. The data model here is based on $\{\xi, \eta\}$ with the size of 1×10. In term of SVD, the matrix A includes all snapshots, $N = 1000$ snapshots and the quantity of sample, S_t was changed, determining the size of training data set. By observing the change of error ϵ through the change of S_t,we can find a reasonable quantity for S_t. Specifically, we only use a tiny part of the data to train the linear regression model, in most validation cases involved in this study.

In contrast, the remaining data points are used to evaluate the training surrogates' generalization ability for other new cases. To compare the accuracy among each linear regression models, the mean and variance of the error estimate are shown in Fig. 7. In both Fig.7 (i) and (ii), the horizontal axis shows the quantity of raw points, S_t, which indirectly limited the quantity of data points used in training the least square model. Since the raw data points is driven from S set of high fidelity data points, which is the most important factor in determining computational expense in this process. Fig.7 (i) shows that the error of the two models are rapidly reduced in the first 13 training points, and the response has been significantly improved. For more than 20 training points, the continuous response speed has been improved.



(i) (ii)

Figure 7. The L_2 mean error (i) and standard deviation (ii) of least square regression with the change of quantity of training points. We construct one hundred linear regression models by randomly sampling the original data set and get the error of L_2 by averaging all the models (Here, the rank of r equal to 3 and T_a related variables ξ and η are two vectors with the size of 5×1.). The low standard deviation indicates the reliability of the model for unobserved cases and the low training points quantity means low computational cost.

The standard deviation of the error can measure the sensitivity to choosing the training points. A high-level standard deviation in error reflects a large variability in the selected training points and, conversely, for the minor standard deviation. In Fig.7 (i) the standard deviations of the errors for different quantities of training points are shown. It can be expected that for a smaller quantity of training points, that is, $N < 15$, the influence relative to the selected point will be greater, while for a larger quantity of training points, the influence of the selected point will be smaller. This behavior can be observed in the Fig.7 (ii).

In the next step, the low-fidelity model will be built based on data set Y

$$f_{LF}^{*} = U_r \Sigma_r Y^{*T} \tag{4-3}$$

where $r = 3$, keeping 94.29% of the matrix A. We want the low fidelity modeling could be computational cheap enough while still providing a rough measurement of the real observation. In Fig.8, the scatter plots of different low-fidelity models and high-fidelity observation are shown. The low-fidelity model are trained with 5 (Fig.8 (i)),11(Fig.8 (ii)) ,15(Fig.8 (iii)),20(Fig.8 (iv)) numerical simulation samples. These figures understandably show that as S_t, the quantity of sample, grow up, the scatter points coalesce around $y = x$ and the variance of the estimate lessen. Thus, we pick $S = 16$ as our low fidelity modeling. Although, the reduced order modeling base on 16 samples of numerical simulation data can not reflect the true observations well, but they are still distributed near the real values. In next section, we can predict the real observations in the multi-fidelity framework and we expect that the results will be significantly improved on the basis of a very small amount of high-precision data.

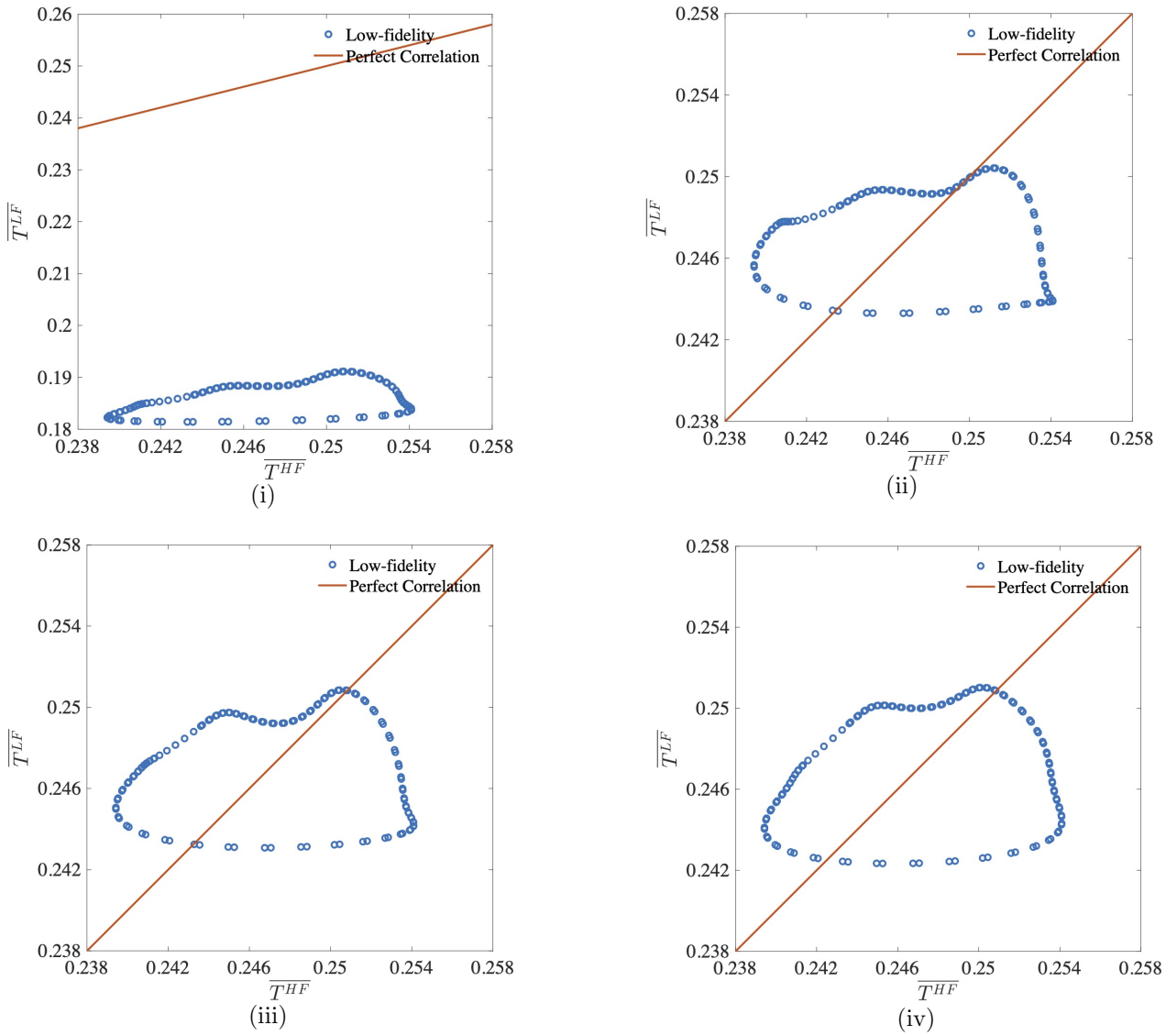


Figure 8. Temporal low-fidelity model: Correlation of average temperature obtained from low-fidelity (f_{LF}) with different quantity of training sets (i)$S_t = 8$, (ii)$S_t = 12$, (iii)$S_t = 16$, (iv)$S_t = 20$. The red line shows a perfect correlation between low–fidelity data and high-fidelity data. These figures show that as the quantity of training sets increases, the low-fidelity data points get closer to the straight line $x = y$.

4.1.2 Temporal Multi-fidelity Modeling

At the beginning of this section, we want to find a suitable data set that is large enough to give a estimation of predictive accuracy for the trained surrogate models. So we can maintain a affordable quantity of high fidelity data points. The multi-fidelity model built here in this section is the one according to the training points, ($N_{LF} = 201$ low-fidelity data points and N_{HF} high-fidelity data points) evenly from $600th$ snapshots to $800th$ snapshots. The quantity of N_{HF} was changed to demonstrate how the quantity of high-fidelity data points effect the accuracy of multi-fidelity modeling. By checking the correlation of average temperature obtained from multi-fidelity and high-fidelity models with different quantity of high-fidelity training points, we want to find a suitable quantity of high-fidelity data in multi-fidelity modeling.

Since the evaluation cost of the low-fidelity model is negligible, we can use the quantity of training points required for the high-fidelity model for functional evaluation as a straight quantify of the computational cost of building the model. In Fig.9, the scatter plots of different multi-fidelity models versus high-fidelity observations are shown. The multi-fidelity models are trained with 5 (Fig.9 (ii)), 9 (Fig.9 (iii)) and 13 (Fig.9 (iv)) high-fidelity points and 201 low-fidelity points. Fig.10 (i) is the scatter plot of low fidelity models versus time, and it is shown here in order to facilitate comparison with other figure. Meanwhile, From Fig.10, the average outflow temperature over time driven from the multi-fidelity model with N_{HF} high-fidelity and 201 low-fidelity model is shown for $\{\xi, \eta\} = 0.5$. Additionally, the uncertainties associated with the predictions can be measured in the multi-fidelity framework.

The grey area shows the $\pm\sigma[\overline{T}]$, standard deviation,around the prediction for the multi-fidelity models.These figures understandably show that with the quantity of high-fidelity points increases, the scatter points merge around $y = x$ and the variance of the estimate decreases.

Here, we hope to be able to establish a suitable data set. This cross-validation set can measure prediction accuracy for trained prediction products while maintaining a reasonable calculation budget. In this study, the method we take is that we divide the data in our validation case into two parts: (1) A small amount of the data is used to train the GP prediction

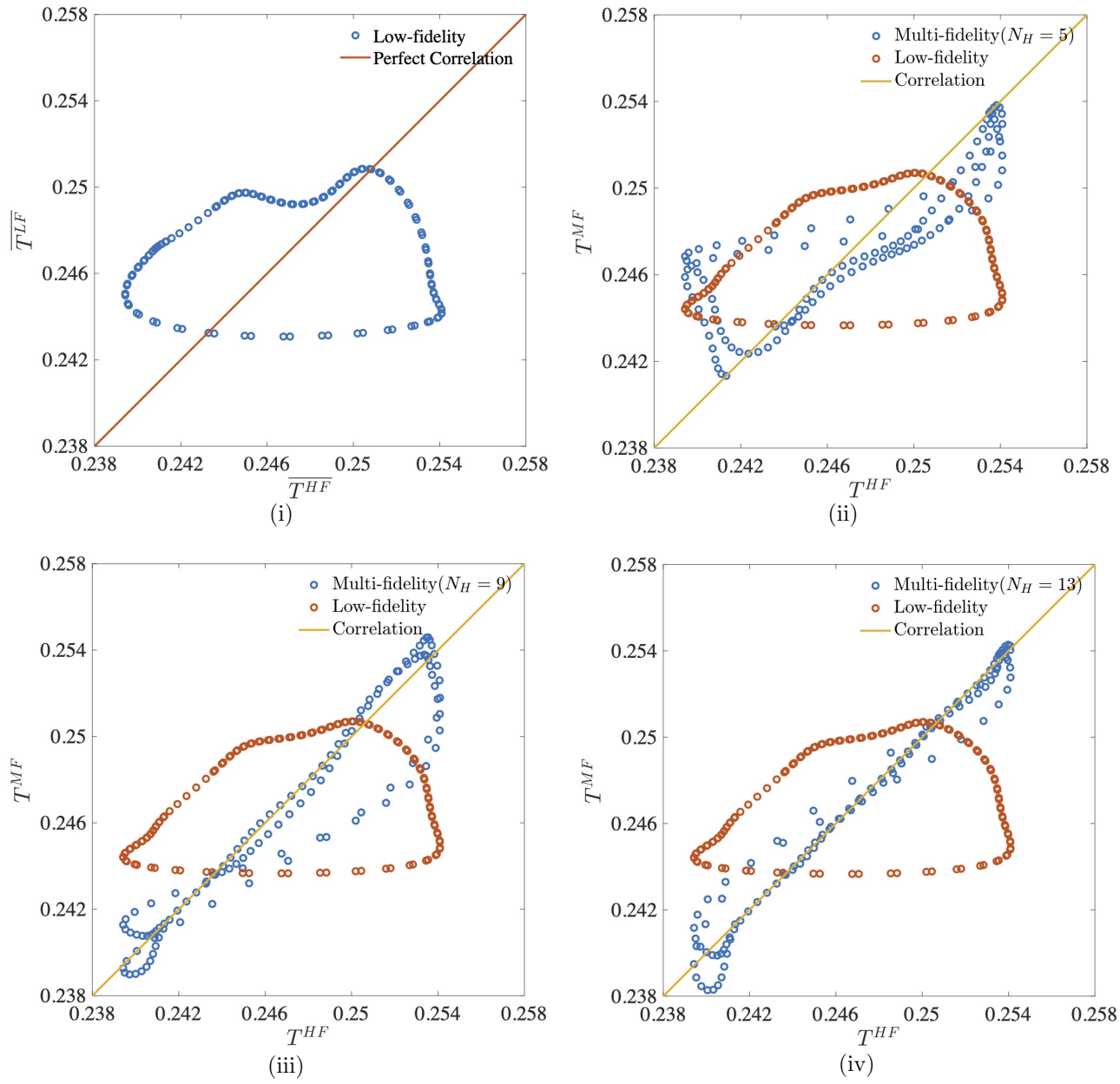


Figure 9. Temporal multi-fidelity model: Correlation of average temperature obtained from (i) low-fidelity and (ii),(iii),(iv) multi-fidelity models with different quantity of high-fidelity data points (ii) $N_{HF} = 5$, (iii) $N_{HF} = 9$, (iv) $N_{HF} = 13$, where 201 low fidelity points are used. These plots show that as the quantity of high-fidelity observations increases, the multi-fidelity model becomes more accurate.

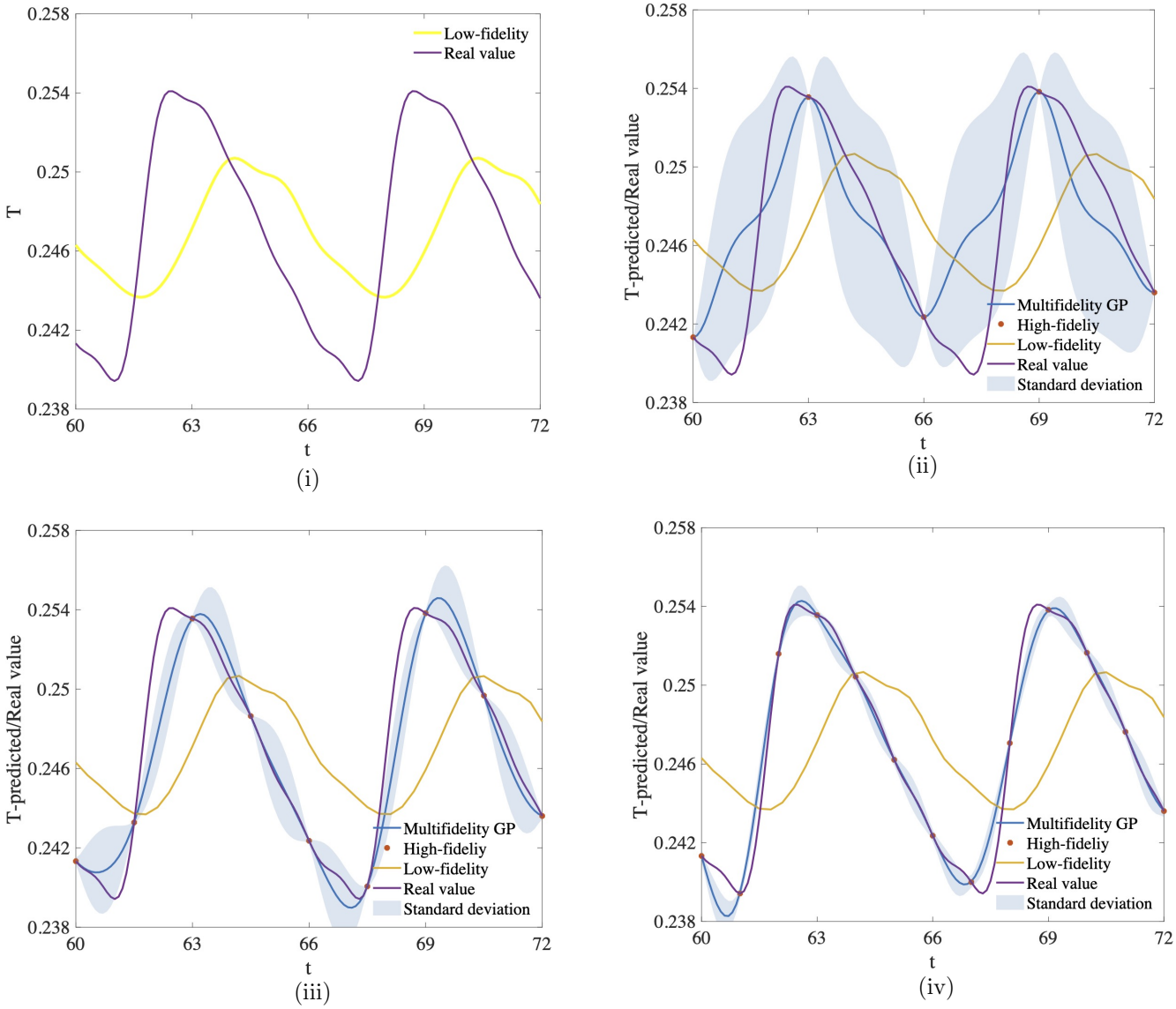


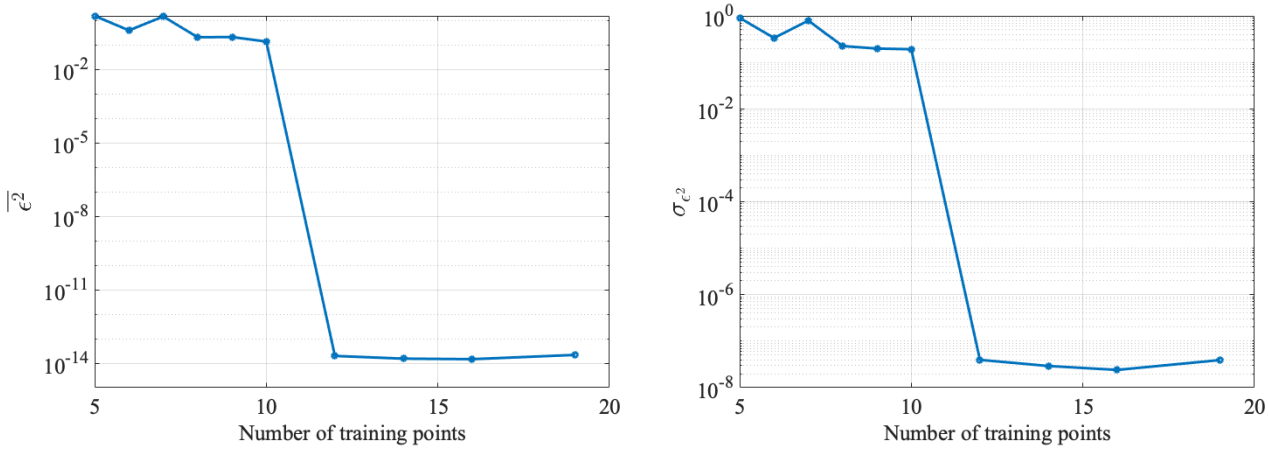
Figure 10. Temporal multi-fidelity model: Average temperature comparisons between multi-fidelity, high-/low-fidelity versus time t. In Fig.10, $N_{LF} = 201$ points are used and (ii), (iii) and (iv) show the effect of improving the quantity of high-fidelity data points, (ii) $N_{HF} = 5$, (iii) $N_{HF} = 9$, (iv) $N_{HF} = 13$. These plots show that as the quantity of high-fidelity observations increases, the multi-fidelity model turns to more accurate.

model (2) We use the remaining data points to evaluate after training The generalization ability of the prediction product for new cases. We want to see the adaptability of the GP model to new, observed data. We use only 201 low-precision samples in this section, which is enough to provide useful information for our GP prediction model to include the characteristics of the temperature distribution on the surface of the elbow's mixed-flow outlet that we have observed. However, we have no worries about computing costs in the sampling of the low-fidelity model. On the other hand, the high-fidelity numerical simulation simulation is performed on the workstation and has a computational cost that cannot be negligible. To this end, five high-fidelity observation data are sufficient to enable our GP prediction model to meet the above verification goals while keeping the computational cost at a manageable level.

4.2 Prediction Result With 2-D Multi-fidelity Modeling

In this section, we want to predict the outflow surface temperature, which means we need to build $2-D$ multi-fidelity framework with data set x, y. Here, $x = \{t, x\}$ contains not only temporal variable t but also space variable x on the outflow surface. The dimension of x will change the construction of matrix A, so the elements in matrix A will be (t, x, ξ, η). We have already discussed in the methodology section. The rest of the settings are the same as in the first $1D$ pedagogical example.

4.2.1 The Cross-validation For Least Square Regression



(i) (ii)

Figure 11. The L_2 mean error (i) and standard deviation (ii) of least square regression with the change of quantity of training points. We construct one hundred linear regression models by randomly sampling the original data set and get the error of L_2 by averaging all the models (Here, the rank of r equal to 3 and T_a related variables ξ and η are two vectors with the size of 5×1.). Compared with the result in previous section, the trends of mean error and standard deviation are similar.

Fig.11 shows that for a small quantity of training points (i.e. $N < 20$), the sensitivity relative to the selected point will be greater, while for larger points, it will be less.

In Fig.12, the horizontal axis shows the temperature value obtained from high-fidelity numerical simulation samples and the vertical axis is the corresponding low-fidelity models are at those points. The scatter plots of different low-fidelity model versus high-fidelity observation are shown. The multi-fidelity models are trained with 10 (Fig.12 (i)), 12 (Fig.12 (ii)), 14 (12 (iii)), 19 (Fig.12 (iv)) training data sets. It is clear from Fig.12 that for the majority of the points the low-fidelity model under-predicts, the temperature is around 0.9, which corresponds to the inlet A side domain in our problem. and when the value of T^{HF} is around 0.1, the points of low-fidelity model is close to the correlation.

In the next step, the low-fidelity model will be built based on data set Y

$$f_{LF}^{*} = U_r \Sigma_r Y^{*T} \tag{4-4}$$

where r = 3, keeping 96.53% of the matrix A. We want the low fidelity modeling could be computational cheap enough while still providing a rough measurement of the real observation. In Fig.12, the scatter plots of different low-fidelity models and high-fidelity observation are shown. The low-fidelity model are trained with different value of S: 10 (Fig.12 (i)),12(Fig.12 (ii)),14(Fig.12 (iii)),19(Fig.12 (iv)) numerical simulation samples. These plot understandably demonstrate that as S, the quantity of sample, increases, the scatter points affiliate around $y = x$ and the variance of the estimate decreases.Thus, we pick $S = 10$ as our low fidelity modeling.

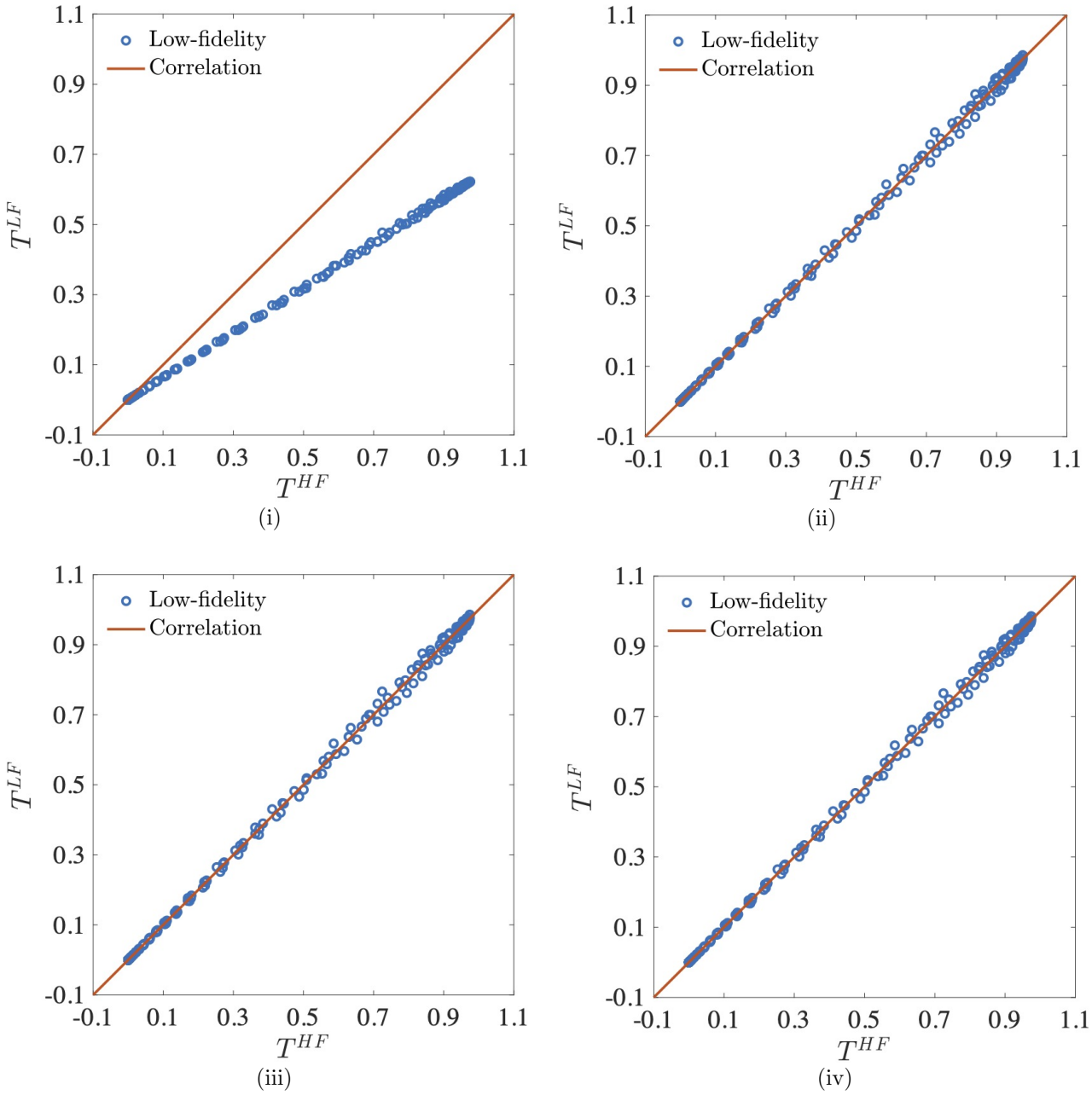


Figure 12. Space-time low-fidelity model: Correlation of temperature obtained from (i),(ii),(iii),(iv) low-fidelity(f_{LF}) with different quantity of training points (i) $N_t = 10$, (ii) $N_t = 12$, (iii)$N_t = 14$, (iv) $N_t = 19$. These plots show that as the quantity of high-fidelity observations increases, the low-fidelity model becomes more accurate.

4.2.2 Space-time Temperature Multi-fidelity Modeling

We hope to find a suitable data set that is large enough to measure the prediction accuracy of the replaced while maintaining an affordable quantity of high-fidelity data points. In this section, the multi-fidelity model constructed here is a model based on some training points, $N_{LF} = 2121$ low fidelity, and in each case, the quantity of N_{HF} will be Changes. By checking the correlation between the outflow surface temperature driven from the multi-fidelity model and the high-fidelity model and different quantities of high-fidelity training points, we hope to find the right amount of high-fidelity data in the multi-fidelity framework.

Since the estimation cost of a low-fidelity model is insignificant, the quantity of training points required for a high-fidelity model for functional evaluation is a straightforward quantify of the computational consuming of building the model. In Fig.13, the scatter plots of different multi-fidelity models versus high-fidelity observations are shown. The multi-fidelity models are trained with 25 (Fig.13 (ii)), 35 (Fig.13 (iii)) and 45 (Fig.13 (iv)) high-fidelity points and 2121 low-fidelity points. Fig.13 (i) is the scatter plot of low fidelity models versus high-fidelity data, and it is demonstrated here for easy juxtaposition with other figures. These figures obviously show that as the number of high-fidelity points increases, the scattered points merge around $y = x$, and the estimated variance decreases. We want to find a suitable quantity of high-fidelity data points for building our multi-fidelity framework, which is larger enough to provide trained surrogates with a measure of prediction accuracy while maintaining a reasonable computational budget.

Then, According to Fig.14, we can analyze the difference between high-precision contour and low-precision contour. In the limited high-fidelity data points, we intentionally select more high-fidelity data randomly at the left end of the outlet C surface and we got our multi-fidelity modeling work in Fig.15.

In Fig.15, our results obtained from multi-fidelity framework in Fig.14 are illustrated by another way. Fig.13 (i), (iii) and (v) show the outflow surface temperature contour over time t and space x. In addition, their corresponding uncertainty are shown in Fig.15 (ii), (iv) and (vi). Meanwhile, Fig.14 shows the outflow surface temperature only obtained from true observation and low-fidelity points. when the quantity of low fidelity points in multi-fidelity

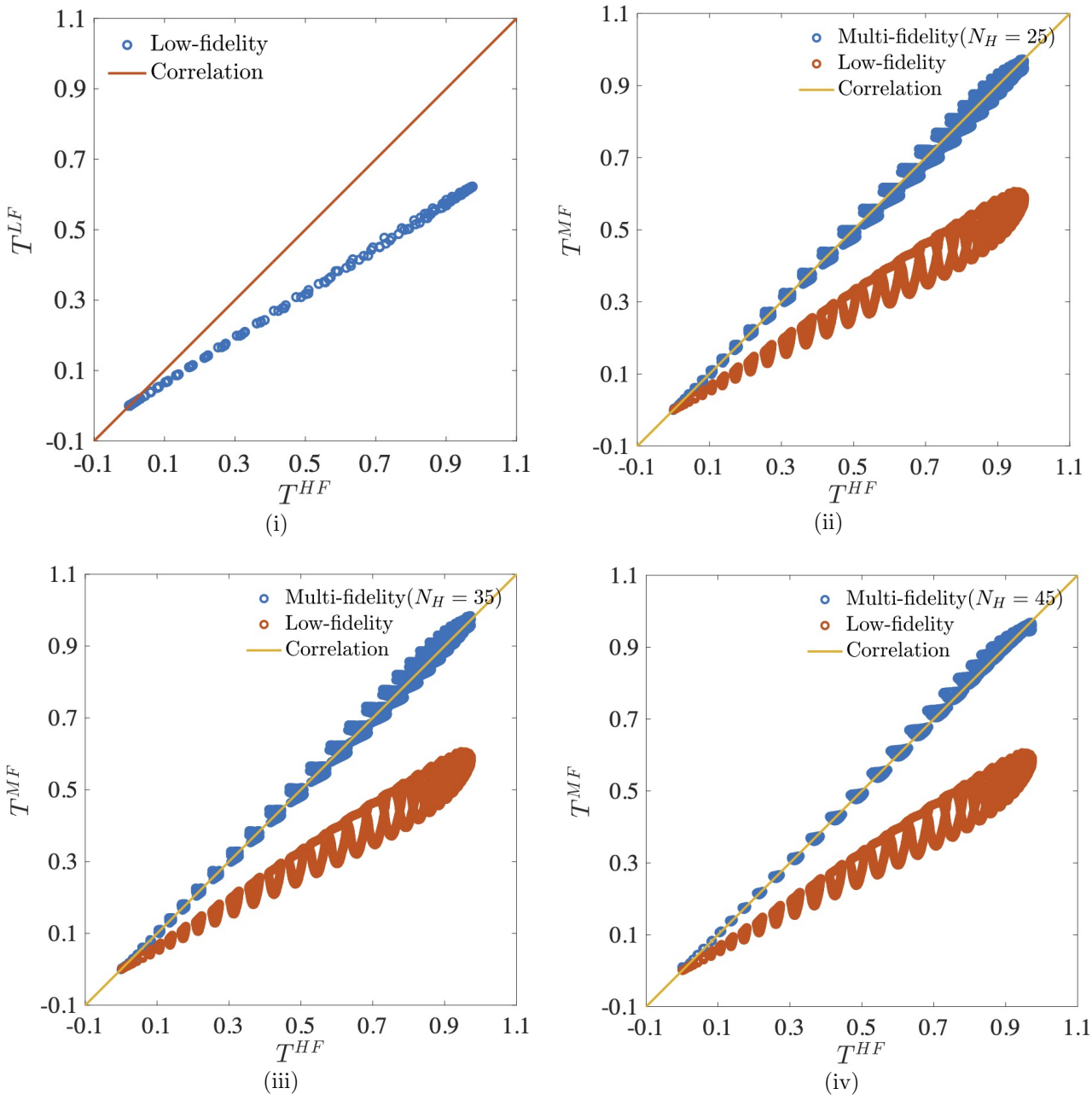


Figure 13. Space-time multi-fidelity model: Correlation of temperature obtained from (i) low-fidelity and (ii),(iii),(iv) multi-fidelity models with different quantity of training points (ii)$N_t = 25$, (iii)$N_t = 35$, (iv)$N_t = 45$. In the multi-fidelity modeling, $N_{LF} = 2121$ points are used. These plots show that as the quantity of high-fidelity data sets grows, the multi-fidelity model becomes more precise.

are constant, depicted in Fig.15 (i),(iii) and (v), as the quantity of high-fidelity data increase, the prediction result on the left side of the exit will be closer to the true value, while the part on the right side has not received much impact. Also, from Fig.15 (ii),(iv) and (vi), we can see the uncertainty of the prediction on the left side of the Fig.is insignificant, this area coincides with the area where the low-fidelity prediction is poor in Fig.14 (ii), which means that a small amount of high-fidelity data can significantly improve the prediction results. In the next chapter, we will compare the prediction results of single-fidelity modeling and multi-fidelity modeling to demonstrate our conclusions further.

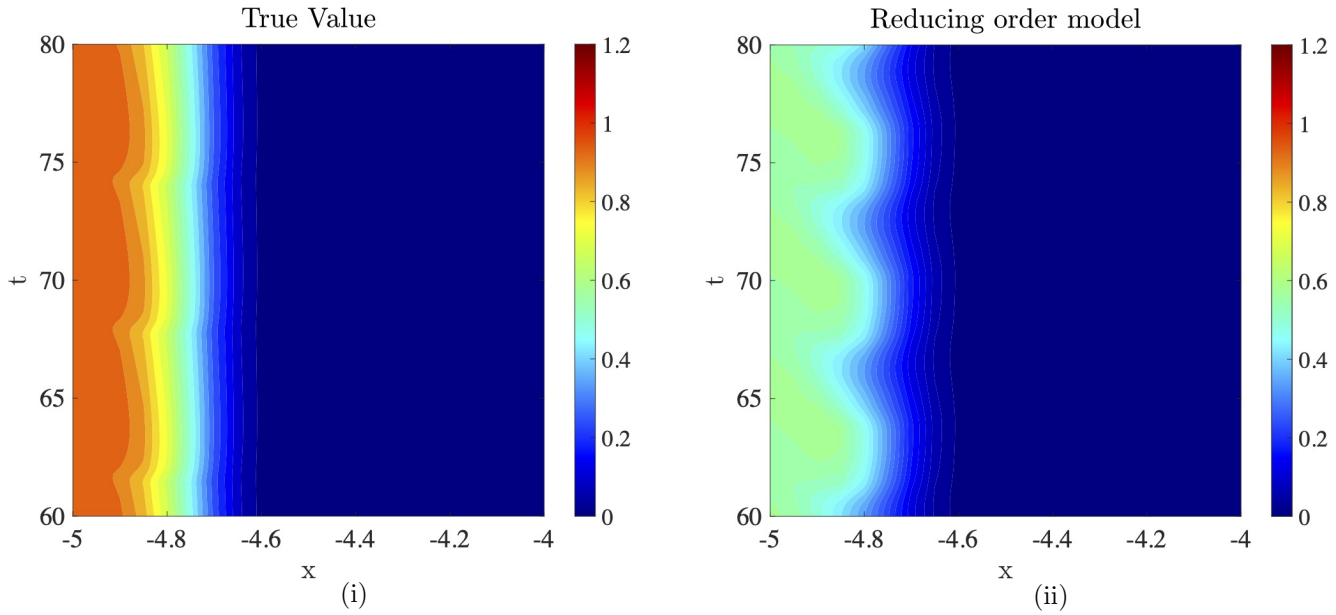


Figure 14. Space-time high-fidelity model and low-fidelity model: the contour for outflow surface temperature over time obtained from true observation and low-fidelity points. Since the temperature change at the right end of outlet C is not obvious, the low-fidelity model is slightly different from it. The high-fidelity data shows that the temperature change period and phase at the left end of outlet C is obviously different from the low-fidelity data.

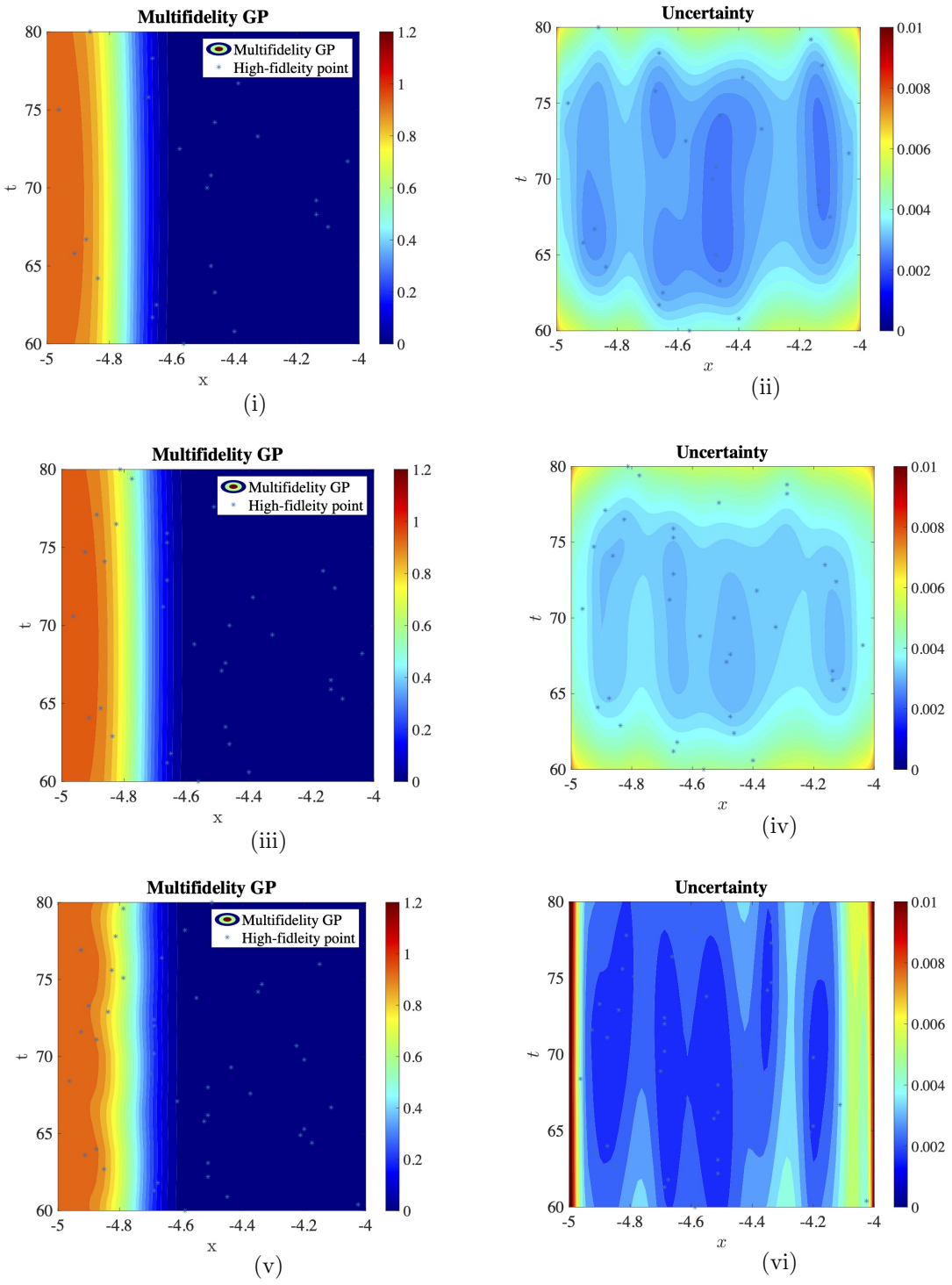


Figure 15. Space-time multi-fidelity model: The contour for outflow surface temperature over time obtained from multi-fidelity models with different quantity of training points, (i)(ii)$N_t = 25$, (iii)(iv)$N_t = 35$, (v)(vi)$N_t = 45$. The Figure of uncertainty shows the $\pm 2\sigma[T]$(standard deviation) around the mean for the multi-fidelity models.

4.2.3 The Comparison With Single-fidelity Modeling

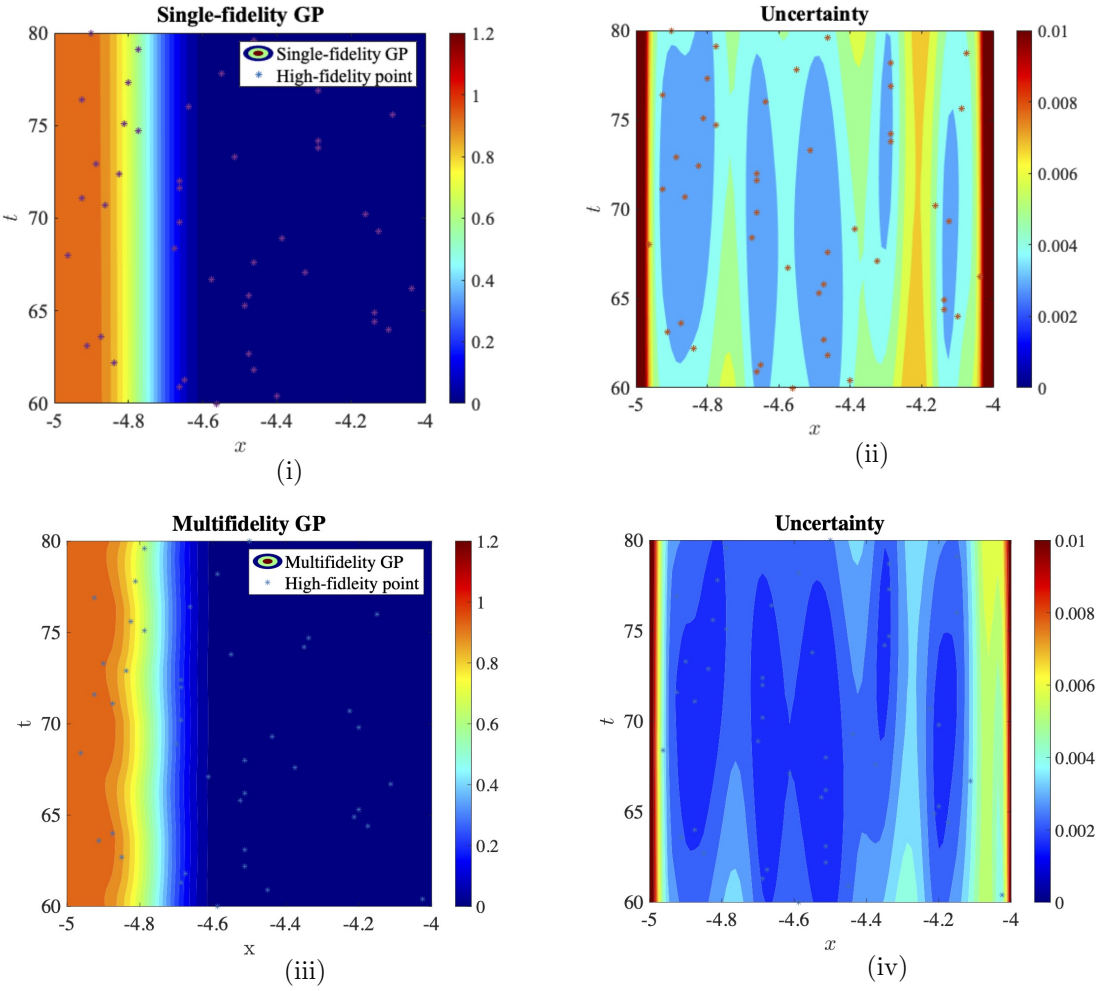


Figure 16. Space-time multi-fidelity model and single-fidelity model: The contour for outflow surface temperature over time obtained from singular-fidelity models and multi-fidelity models; the quantity of high-fidelity data points are both 45. Obviously, a large amount of low-precision data points has greatly improved the accuracy of the multi-fidelity model.

Here, we compare the results of our multi-fidelity model and single-fidelity model. We use 45 high-fidelity data points for constructing the single-fidelity model; meanwhile, we combine these 45 high-fidelity points with 2121 low-fidelity points in our multi-fidelity framework. The results are shown in the Fig.16.

From Fig.16 (ii) and (iv), the distribution of uncertainty values is similar and the value in Fig.(iv) is slightly higher than the value in Fig.16 (ii). However, when we compare Fig.16 (iii) and (i) with Fig.14(i), Multi-fidelity modeling can more accurately predict temperature changes in time and space, which is not reflected in the results of singular-fidelity modeling. Therefore, a large quantity of sufficiently complex and inexpensive low-fidelity data points can enable multi-precision modeling to significantly improve its prediction accuracy based on a small amount of high-precision data.

5.0 Conclusions

In this work, we introduced a multi-fidelity framework based on modern machine learning technology. We obtained alternative models that can be quickly evaluated by combining low-fidelity and high-fidelity data. This framework compensates the lack of sufficient high-fidelity data by combining the scarce accurate measurements with cheap to evaluate low-fidelity measurements. The presented framework combines these two different levels of data in a Bayesian framework where the uncertainty of each level of fidelity is accounted for in a principled manner. In particular, multi-fidelity Gaussian process (GP) regression is employed, where a GP is learned at each levels of fidelity and moreover, the correlations between two levels of fidelity is also learned. Because of the Bayesian treatment of of the regression, the predictions are endowed with uncertainties. These uncertainties are the result of having insufficient data as well noise and other uncertainties in each level of fidelity. The uncertainties can be used to guide further data gathering by building various acquisition functions.

Despite enjoying many favorable features, the GP regression has one major challenge: training the GP scales with $O(N^3)$, where N is the size of the training data points (low-fidelity as well high-fidelity training pints). For high-dimensional regression problems, where the size of training data needs to grow – often exponentially with the dimension of the input space – this training cost could be unaffordable. In this work, we address this challenge by performing dimension reduction of the input space and then build the GP using the reduced space.

We have demonstrated the performance of the proposed strategy for transient heat transfer problem with various types of quantities of interest. With the development of machine learning tools, the possible direction of future work is to use deep and recursive Bayesian learning techniques to infer such non-linear graphs from data directly.(see [4]; [13]). Besides, the presented framework can be extended to diverse set of science and engineering problems.

Bibliography

[1] F. Aulery, A. Toutant, R. Monod, G. Brillant, and F. Bataille. Numerical simulations of sodium mixing in a t-junction. *Applied Thermal Engineering*, 37:38–43, 2012.

[2] H. Babaee, P. Perdikaris, C. Chryssostomidis, and G. E. Karniadakis. Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations. *Journal of Fluid Mechanics*, 809:895–917, 2016.

[3] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[4] Andreas C. Damianou and Neil D. Lawrence. Deep gaussian processes, 2013.

[5] Moustafa El-Gammal, H. Mazhar, J. Cotton, C. Shefski, John Pietralik, and C. Ching. The hydrodynamic effects of single-phase flow on flow accelerated corrosion in a 90-degree elbow. *Nuclear Engineering and Design - NUCL ENG DES*, 240:1589–1598, 06 2010.

[6] Alexander Forrester, Andras Sobester, and Andy Keane. Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc. A*, 463:3251–3269, 12 2007.

[7] G.H. Golub and C. Reinsch. *Singular Value Decomposition and Least Squares Solutions*, volume 14, pages 134–151. 01 1971.

[8] S.T. Jayaraju, E.M.J. Komen, and Emilio Baglietto. Suitability of wall-functions in large eddy simulation for thermal fatigue in a t-junction. *Nuclear Engineering and Design*, 240:2544–2554, 10 2010.

[9] G. Karniadakis and S. Sherwin. *Spectral/hp Element Methods for Computational Fluid Dynamics: Second Edition*. Numerical Mathematics and Scientific Computation. OUP Oxford, 2013.

[10] MC Kennedy and A O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 03 2000.

[11] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. Theses, Université Paris-Diderot - Paris VII, October 2013.

[12] Loic Le Gratiet and Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. 4:365–386, 01 2014.

[13] César Lincoln C. Mattos, Zhenwen Dai, Andreas Damianou, Jeremy Forth, Guilherme A. Barreto, and Neil D. Lawrence. Recurrent gaussian processes, 2016.

[14] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160751, 2017.

[15] Paris Perdikaris, D. Venturi, Johannes Royset, and George Karniadakis. Multi-fidelity modelling via recursive co-kriging and gaussian–markov random fields. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 471:20150018, 07 2015.

[16] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Machine learning of linear differential equations using gaussian processes. *Journal of Computational Physics*, 348:683–693, Nov 2017.

[17] CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, January 2006.

[18] Simo Särkkä. Linear operators and stochastic partial differential equations in gaussian process regression. In *International Conference on Artificial Neural Networks*, pages 151–158. Springer, 2011.

[19] Linda M. Sroka and Larry J. Forney. Fluid mixing in a 90.degree. pipeline elbow. *Industrial & Engineering Chemistry Research*, 28(6):850–856, 1989.